DOCUMENT RESUME

ED 475 047 EC 309 497

AUTHOR	Thompson, Sandra; Blount, Amanda; Thurlow, Martha
TITLE	A Summary of Research on the Effects of Test Accommodations: 1999 through 2001. Technical Report.
INSTITUTION	National Center on Educational Outcomes, Minneapolis, MN.; Council of Chief State School Officers, Washington, DC.; National Association of State Directors of Special Education, Alexandria, VA.
SPONS AGENCY	Special Education Programs (ED/OSERS), Washington, DC.
REPORT NO	NCEO-34
PUB DATE	2002-12-00
NOTE	67p.
CONTRACT	H326G000001
AVAILABLE FROM	National Center on Educational Outcomes, Univ. of Minnesota, 350 Elliott Hall, 75 East River Rd., Minneapolis, MN 55455 ($15). Tel: 612-624-8561; Fax: 612-624-0879; e-mail: nceo@umn.edu. For full text: http://www.education.umn.edu/NCEO.
PUB TYPE	Reports - Research (143)
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	Computer Assisted Testing; *Disabilities; Educational Testing; Elementary Secondary Education; Performance Factors; Reading Aloud to Others; Research Design; Research Needs; *Scores; *Test Format; Test Results; *Testing Accommodations; Timed Tests
IDENTIFIERS	Oral Examinations

ABSTRACT

This report updates and summarizes what is known from research on the effects of test accommodations for students with disabilities and also provides direction to the design of critically needed future research on accommodations. A review was conducted of 46 empirical research studies on accommodations published from 1999 through 2001. The majority of the studies involved criterion referenced tests used for state accountability. Findings across at least four studies from the review indicate three accommodations showed a positive effect on student test scores: computer administration; oral presentation; and extended time. However, additional studies on each of these accommodations also found no significant effect on scores or alternations in item comparability. All of the meta analyses of accommodated conditions found a positive effect on scores, and all of the studies examining differential item functioning (DIF) under accommodated conditions found some items that exhibited DIF. The analysis also indicates a need for the clear definition of the constructs tested, greater clarity in the accommodations needed by individual students, and exploration of the desirability and perceived usefulness of accommodations by students themselves. Several appendices include a summary of the types of accommodations and research results. (Contains 54 references.) (CR)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Technical Report 34

ED 475 047

A Summary of Research on the Effects of Test Accommodations: 1999 through 2001

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Thurlow

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

NATIONAL CENTER ON EDUCATIONAL OUTCOMES

In collaboration with:

Council of Chief State School Officers (CCSSO)

National Association of State Directors of Special Education (NASDSE)

EC 309497

2 BEST COPY AVAILABLE

Technical Report 34

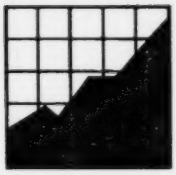
A Summary of Research on the Effects of Test Accommodations: 1999 through 2001

Sandra Thompson • Amanda Blount • Martha Thurlow

December 2002

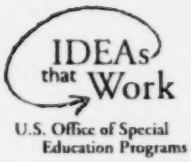
All rights reserved. Any or all portions of this document may be reproduced and distributed without prior permission, provided the source is cited as:

Thompson, S., Blount, A., & Thurlow, M. (2002). *A summary of research on the effects of test accommodations: 1999 through 2001* (Technical Report 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.



NATIONAL CENTER ON EDUCATIONAL OUTCOMES

The Center is supported through a Cooperative Agreement (#H326G000001) with the Research to Practice Division, Office of Special Education Programs, U.S. Department of Education. The Center is affiliated with the Institute on Community Integration at the College of Education and Human Development, University of Minnesota. Opinions expressed herein do not necessarily reflect those of the U.S. Department of Education or Offices within it.



NCEO Core Staff

Deb A. Albus
Jane L. Krentz
Kristi K. Liu
Jane E. Minnema
Michael L. Moore
Rachel F. Quenemoen
Dorene L. Scott
Sandra J. Thompson
Martha L. Thurlow, Director

Additional copies of this document may be ordered for $20.00 from:

National Center on Educational Outcomes
University of Minnesota • 350 Elliott Hall
75 East River Road • Minneapolis, MN 55455
Phone 612/624-8561 • Fax 612/624-0879
http://education.umn.edu/NCEO

The University of Minnesota is committed to the policy that all persons shall have equal access to its programs, facilities, and employment without regard to race, color, creed, religion, national origin, sex, age, marital status, disability, public assistance status, veteran status, or sexual orientation.

This document is available in alternative formats upon request.

Executive Summary

The enactment of the No Child Left Behind Act of 2001 brings an urgency to knowing whether the use of certain accommodations threatens test validity or score comparability, and whether specific accommodations are useful for individual students. This report is intended to update and summarize what we know from research on the effects of accommodations, and also to provide direction to the design of critically needed future research on accommodations. We found 46 empirical research studies on accommodations published from 1999 through 2001. These studies had the following characteristics:

Purpose. The primary purpose of the 1999-2001 accommodations research was to determine the effects of accommodations use on the large-scale test scores of students with disabilities.

Types of assessment, content areas, and accommodations. The majority of the studies involved criterion referenced tests used for state accountability. Mathematics was assessed in half of the studies, and reading/language arts was assessed in about one third. Presentation accommodations were investigated most frequently, with "oral presentation" selected for analysis in nearly half of the studies.

Participants. The number of participants ranged from 3 to nearly 21,000. The largest number of studies included elementary school students, with the greatest number examining accommodation use by fourth graders. Twenty-seven studies documented the participants' types of disabilities; in those studies, learning and cognitive disabilities were most frequently investigated.

Research design. The studies were identified as representing one of four group research designs, a single subject research design, or a non-experimental or other design. Over one third of the studies applied non-experimental or other designs to the study of accommodations effects.

Findings. Despite the variability in the characteristics of the accommodations research conducted from 1999-2001, the findings point to further directions for research. In terms of results, three accommodations showed a positive effect on student test scores across at least four studies: computer administration, oral presentation, and extended time. However, additional studies on each of these accommodations also found no significant effect on scores or alterations in item comparability. All of the meta analyses of accommodated conditions found a positive effect on scores, and all of the studies examining differential item functioning (DIF) under accommodated conditions found some items that exhibited DIF.

Limitations. The researchers of the accommodations studies often identified limitations

in their studies. These are important limitations that need to be given more attention in the future. Among the frequently cited limitations were: unknown variations among students included in the study, sample sizes too small to provide adequate statistical support, and nonstandard administration of the accommodations across proctors and schools. These limitations and other considerations led researchers to recommend replicating the research for validation and generalization, as well as investigating associations to specific disabilities. It was also recommended that more detailed non-experimental studies be conducted to provide richer data, increasing researcher control of the testing process, and studying larger groups of students.

Important overall observations from our analysis include a need for the clear definition of the constructs tested, greater clarity in the accommodations needed by individual students, and exploration of the desirability and perceived usefulness of accommodations by students themselves—the "end users" of assessments. Future research should also explore the effects of assessment design and standardization to see whether incorporating new item designs and incorporating more flexible testing conditions (i.e., universal design) reduces the need for accommodations while facilitating measurement of the critical constructs for students with disabilities.

Table of Contents

Overview 1
Method 2
Results 3
- Purpose of the Research 3
- Type of Assessment 4
- Content Area Assessed 5
- Type of Accommodation 5
- Research Participants 6
- Research Design 8
- Research Results 10
- Limitations Cited by Researchers 12
- Recommendations for Future Research 13

Discussion and Implications for Future Research 14
References 18
Appendix A: Summary of Research Purpose 25
Appendix B: Summary of Type of Assessment 31
Appendix C: Summary of Content Area Assessed 35
Appendix D: Summary of Types of Accommodations 37
- Specifications for Appendix D 41

Appendix E: Summary of Participants 43
Appendix F: Summary of Research Designs 47
Appendix G: Summary of Research Results 49
Appendix H: Summary of Limitations Cited by Researchers 55
Appendix I: Summary Recommendations by Researchers 59

Overview

One of the most viable ways to increase the participation of students with disabilities in assessments is through the use of accommodations. Accommodations are defined by Thurlow and Bolt (2001) as "changes in assessment materials or procedures that address aspects of students' disabilities that may interfere with the demonstration of their knowledge and skills on standardized tests. Accommodations attempt to eliminate barriers to meaningful testing, thereby allowing for the participation of students with disabilities in state and district assessments" (p. 1). Accommodations are further defined by Tindal and Fuchs as "changes in standardized assessment conditions introduced to level the playing field for students by removing the construct-irrelevant variance created by their disabilities. Valid accommodations produce scores for students with disabilities that measure the same attributes as standard assessments measured in nondisabled individuals" (p. 8).

All states now have accommodation policies (Thurlow, Lazarus, Thompson, & Robey, 2002), and nearly sixty percent of states keep track of the use of accommodations during state assessments—about half of these report an increase in use by students (Thompson & Thurlow, 1999). However, there continues to be only limited consensus on what constitutes an "appropriate" accommodation as states grapple with decisions about how to score and report the use of accommodations that some consider "nonstandard" or "nonscorable."

There is a critical and ongoing need for increased research on the effects of the use of accommodations on the psychometric characteristics of assessment results. With the enactment of the No Child Left Behind Act of 2001 has come an urgency to know whether the use of certain accommodations threatens test validity or score comparability. Similarly, there is a need to know whether specific accommodations are useful for individual students (Thurlow, McGrew, Tindal, Thompson, Ysseldyke, & Elliott, 2000). The amount of assessment research has increased dramatically in recent years. Tindal and Fuchs (1999) found an increase from 11 studies published from 1990 through 1992, to 29 studies published from 1996 through 1998. The current study, which focuses on studies published from 1999 through 2001, includes 46 empirical research studies on accommodations (see Table 1).

Table 1. Number of Accommodations Research Studies Published from 1990 Through 2001

Years	Number of Studies
1990 through 1992	11
1993 through 1995	18
1996 through 1998	29
1999 through 2001	46

This increase is partly due to support from the U.S. Department of Education for research on a variety of issues related to the participation of students with disabilities in large-scale assessments. Federal support has come from both the Office of Special Education Programs (OSEP) and the Office of Educational Research and Improvement (OERI). Some states and test publishers have also recently supported additional research efforts.

The purpose of this paper is to summarize several components of the research on the effects of test accommodations published from 1999 through 2001, including: type of assessment, content area assessed, number of research participants, types of disabilities included in the sample, grade-level of the participants, research design, research findings, limitations of the study, and recommendations for future research.

Method

Four major databases were searched to identify research on test accommodations published from 1999 through 2001: ERIC, PsychInfo, Educational Abstracts, and Digital Dissertations. Research papers were also obtained at major conferences. Additional resources for identifying research included:

- Behavioral Research and Teaching at the University of Oregon: http://brt.uoregon.edu/
- Education Policy Analysis Archives: http://epaa.asu.edu
- National Center for Research on Evaluation, Standards, and Student Testing: http://www.cse.ucla.edu/
- Wisconsin Center for Educational Research: http://www.wcer.wisc.edu/testacc/

Several search terms were used. The terms were varied systematically to ensure the identification of all research on changes in testing, published from 1999 through 2001. Search terms included:

- accommodation
- test adaptation
- test changes
- test modifications

- test accommodations
- state testing accommodations
- standards-based testing accommodations
- large-scale testing accommodations

A decision was made to limit the selection of publications to empirical research. Included within this realm are studies with samples consisting of preschool, kindergarten through high school, and postsecondary students. The focus of the empirical research was not limited only to large-scale testing, but also included studies that incorporated intelligence tests and curriculum-based measures (CBM). We decided to focus on testing accommodations as opposed to instructional accommodations, although there is some overlap between these purposes in the literature. We did not include any conceptual or opinion pieces in this analysis.

Results

As a result of the extensive search effort described above, 46 research studies, published between 1999 and 2001, were selected for this analysis. All of the studies are empirical, that is, they include an analysis of data. Nineteen of the studies were published in journals, 12 in reports, 12 in papers presented at conferences, and 3 were dissertations. The researchers and references to each publication are listed in Appendix A.

Purpose of the Research

The primary purpose of the accommodations research conducted over the past three years has been to determine the effect of accommodations use on the large-scale test scores of students with disabilities (see Table 2). Over half of the studies investigated whether the use of accommodations gave the test scores of students with disabilities a differential boost, that is, the accommodation had a greater effect on the scores of students with disabilities than on the scores of students without disabilities. The second most common purpose was to investigate the effects of accommodations on test score validity. The purpose of seven of the studies was to analyze institutional factors, teacher judgment, or student desirability of accommodation use. Three of these studies also examined the effect of accommodation use on test scores. Finally, five studies described as their purpose the examination of patterns of errors across items or tests. Though we have categorized the purposes into only four primary groups, the results of this analysis show wide variation in research designs, participants, and assessments. Appendix A summarizes the purpose of each study.

Table 2. Purposes of Reviewed Research

Research Purpose	Number of Studies
Determine the effect of the use of accommodations on test scores of students with disabilities	24
Investigate the effects of accommodations on test score validity	10
Study institutional factors, teacher judgment, or student desirability of accommodation use	7
Examine patterns of errors across items or tests	5

Type of Assessment

Three primary types of assessments were used across the 46 studies selected for this analysis. These included norm-referenced and other standardized tests, state criterion-referenced tests or performance assessments, and school or district-designed tests (see Table 3). Three of the studies used assessments from more than one category. Seventeen studies used a total of 10 different norm-referenced and other standardized tests including: National Assessment of Educational Progress, Test of General Educational Development, Stanford Achievement Test, Iowa Test of Basic Skills, Peabody Picture Vocabulary Test, California Achievement Test, Terra Nova, Scholastic Ability Test for Adults, Psychoeducational Profile, and Miller Analogies Test.

Studies using criterion-referenced tests consisted primarily of large-scale state accountability measures. These assessments, used in 21 of the studies selected for this analysis, crossed a total of 13 states: Indiana, Kansas, Kentucky, Massachusetts, Maryland, Minnesota, Missouri, New York, Oregon, Rhode Island, South Carolina, Washington, and Wisconsin. Four studies were conducted in each of two states: Oregon and Wisconsin.

Six states used school or district-designed tests. These included performance assessments, curriculum-based measures, and math computation tests. The "other" category consisted of three surveys, a checklist, and two studies that included an analysis of multiple investigations (e.g., meta analysis). Appendix B lists the type of assessments used in each study.

Table 3. Types of Assessment in Reviewed Research

Type of Assessment	Number of Studies*
Norm-referenced and Other Standardized Tests	17
State Criterion-referenced Tests or Performance Assessments	21
School or District-designed Tests	6
Other	6

* Some studies had assessments that fit into more than one category.

Content Area Assessed

Researchers used assessments across five basic academic content areas: reading/language arts, writing, mathematics, science, and social studies. Mathematics was assessed in half of the studies, while reading/language arts was assessed in 16 studies (see Table 4). The studies categorized as "no specific content area" included surveys, meta analyses of several studies, and general academic assessments in which specific content was not specified. Most of the research focused on a single content area, while some of the larger studies addressed two or more content areas (see Appendix C).

Table 4. Content Areas Assessed in Reviewed Research

Content Areas Assessed	Number of Studies*
Mathematics	23
Reading/Language Arts	16
Science	9
Writing	7
Social Studies	3
No Specific Content Area	9

* Some studies assessed more than one content area.

Type of Accommodation

Eleven types of accommodations were investigated in at least two of the 46 research studies selected for this analysis (see Table 5). These accommodations were categorized into four groups: presentation, response, setting, and timing/scheduling (see Appendix D). In addition, 14 of the studies investigated the effects of multiple accommodations.

Presentation accommodations were investigated most frequently, with "oral presentation" examined in 22 studies. Other presentation accommodations included computer administration, simplified language, and large print. Timing and scheduling accommodations were the next most frequently investigated, with extended time analyzed in 17 studies. Other timing and scheduling accommodations studied included testing over multiple days and the use of frequent breaks. Response accommodations followed and included dictated response, use of a word processor, and calculator use. Finally, use of a separate or individual setting was investigated in five studies.

The 19 studies listed in the "other" category include accommodations investigated in only one study. These included: extra spacing, assistive devices, repeated directions, verbal encouragement, cueing, interpreter, and support in understanding directions. In addition, survey research not aimed at investigating the effects of a particular accommodation was recorded as "other."

Table 5. Types of Accommodation in Reviewed Research

Type of Accommodation		Number of Studies*
Presentation:	Oral Administration	22
	Computer Administration	8
	Simplified Language	6
	Large Print	2
Response:	Dictated Response	6
	Word Processor	4
	Calculator	2
Setting:	Separate setting/small group	5
Timing/Scheduling:	Extended Time	17
	Multiple Day	2
	Frequent Breaks	2
Multiple Accommodations		14
Other		19

* Some studies assessed more than one accommodation.

Research Participants

A description of the research participants, including the number in each study, percent of participants with disabilities, grade level or age, and types of disabilities are described in Appendix E. The number of participants ranged from 3 to nearly 21,000. Table 6 shows that about half of the studies included fewer than 200 participants. The largest study included 20,791 participants in Kentucky's statewide assessment over a two-year period (Koretz & Hamilton, 2000).

Table 6. Number of Participants in Reviewed Research

Number of Participants	Number of Studies
1 – 99	11
100 – 199	12
200 – 299	6
300 – 499	4
500 – 999	2
More than 1000	7
Number Unknown	4

Thirty-two studies documented the number of participants with disabilities (see Table 7). The percent of students with disabilities ranged from 6 percent to 100 percent of the total sample, with 7 studies including fewer than 25% students with disabilities, 12 fewer than 50%, 8 fewer than 75%, and 5 studies including 75 to 100 percent of those with disabilities. Eight studies

included students with disabilities, but the number or percent was not documented. Six studies were reviews or other approaches that did not directly include participants.

Table 7. Percent of Sample Consisting of Students with Disabilities in Reviewed Research

Percent of Sample Consisting of Students with Disabilities	Number of Studies
1 – 24%	7
25 – 49%	12
50 – 74%	8
75 – 100%	5
Percent Unknown	8
Not Applicable	6

Participants in the research studies ranged in age from elementary school through postsecondary education (see Table 8). The largest number of studies (16) included elementary school students, with the greatest number of studies examining accommodation use by fourth graders. Eleven studies examined accommodation use by middle school students and six studies included high school students. Six studies looked at accommodation use across all grade levels, and three studied postsecondary students. Grade level information did not apply to or was not reported in four studies.

Table 8. Summary of Participant Grade Levels in Reviewed Research

Participant Grade Level	Number of Studies
Elementary (grades k-5)	16
Middle School (grades 6-8)	11
High School (grades 9-12)	6
Multiple Grade Levels (k-12)	6
Postsecondary	3
Not Applicable	4

Twenty-seven studies documented the types of disabilities experienced by participants (see Table 9). Some of the studies included students with a variety of disabilities, while others focused on a single disability (e.g., learning disability) or deficit area (e.g., reading). Three studies included students representing all disability categories. Other than these three studies, only one included students with hearing impairments. Students with visual or physical disabilities were not included (see Appendix E).

Table 9. Types of Disabilities Experienced by Participants in Reviewed Research

Type of Disability	Number of Studies
Learning Disability	18
Cognitive Disability (e.g., mental retardation)	10
Emotional/behavioral Disability	9
Communication Disability	7
Reading or Math Deficit	5
Other (includes physical and sensory disabilities, autism, attention deficit disorder, health impairments, and multiple disabilities)	9

Research Design

The research designs used in the studies selected for this analysis (see Appendix F) were organized according to types identified by Thurlow et al. (2000). These included four types of group research designs, a single subject research design, and a non-experimental or other design. The four group research designs are shown in Figure 1.

Figure 1. Group Research Designs

Design 1: Score comparability, interaction between presence of disability and accommodation use, equivalent test forms

	Disability Group 1	Disability Group 2	Non-Disability Group 1	Non-Disability Group 2
With Accommodation	Test Form A	Test Form B	Test Form A	Test Form B
Without Accommodation	Test Form B	Test Form A	Test Form B	Test Form A

Design 2: Score comparability, interaction between presence of disability and accommodation use, matched sample

	Disability Group 1	Disability Group 2	Non-Disability Group 1	Non-Disability Group 2
With Accommodation	Test Form A		Test Form A	
Without Accommodation		Test Form A		Test Form A

Design 3: Score comparability with accommodation use, equivalent test forms

	Disability Group 1	Disability Group 2
With Accommodation	Test Form A	Test Form B
Without Accommodation	Test Form B	Test Form A

Design 4: Score comparability with accommodation use, matched sample

	Disability Group 1	Disability Group 2
With Accommodation	Test Form A	
Without Accommodation		Test Form A

The four group research designs shown in Figure 1 differ in terms of the controls that are included and the requirements for matching of the samples. For example, in Design 1 participants take two equivalent forms (A and B) of the same test – one with and the other without accommodations. Participants with and without disabilities who take the test without the accommodations are drawn from the general testing population. Their scores are randomly selected from the total test sample of all students who regularly take the version of Forms A and B. This design does not require that the sample from the disability and non-disability groups be exactly similar (i.e., matched) in important characteristics.

In Design 2, only one form of the test is used, but students in the disability groups and the non-disability groups must be matched samples that are equivalent in important characteristics (e.g., age, disability category, accommodation need, etc.). If the students are not matched on important characteristics, it is impossible to determine whether any differences between the score characteristics of the groups are due to the effects of the accommodations or are attributable to differences in sample characteristics. If appropriate matching can take place, it is possible for participants with and without disabilities who take the test without accommodations to be drawn from the general testing population.

In Design 3, score comparability as a function of accommodation use is examined only for students with disabilities. This design assumes (based on prior research) that the scores of participants who take the test without the accommodation are comparable to the scores of participants without disabilities who take the test without accommodations. Participants must take two versions of the same test – one with and one without accommodations. It is possible to draw participants who take the test without accommodations from the general testing population taking both Form A and Form B.

Design 4 also examines score comparability only with participants with disabilities. The design, like Design 3, assumes (based on prior research) that the scores of students with disabilities who take the test without the accommodation are comparable to students without disabilities who take the test without the accommodation. This design requires that the participants be matched samples, but does not require equivalent forms (thereby allowing the use of just one test form). Without matched samples, it is impossible to determine whether differences in score characteristics are due to the effect of the accommodation or to differences in sample characteristics.

Table 10 shows the number of studies that used each of the four group designs, as well as the single subject and non-experimental and other designs. As is evident in this table, 12 of the accommodations studies in the past three years used Design 1, by far the most of any of the designs other than the non-experimental and other. The three studies that used single subject designs generally were intended to determine whether an accommodation is effective for individual students, and perhaps to search out the reason for the effects. These studies monitored student performance over time, along with the systematic introduction of various "treatments" that are considered to be accommodations. The 17 studies that were in the last category included a variety of methods such as meta-analyses, survey research, investigations of differential item functioning, post-hoc comparisons of scores, and methods for testing the fit of various models. A complete summary of the research designs used across all studies is provided in Appendix F.

Table 10. Research Designs in Reviewed Research

Type of Research Design	Number of Studies*
Group Research Design 1: Score comparability, interaction between presence of disability and accommodation use, equivalent test forms	12
Group Research Design 2: Score comparability, interaction between presence of disability and accommodation use, matched sample	2
Group Research Design 3: Score comparability with accommodation use, equivalent test forms	7
Group Research Design 4: Score comparability with accommodation use, matched sample	6
Single Subject Research Design	3
Non-experimental/Other	17

* All studies except one fit within only one of the categories. The one study in which multiple categories was coded as both Design 1 and Design 3 because not enough information was available to distinguish between the two.

Research Results

Results of the primary accommodations studied in the 46 research studies that we reviewed,

using the designs described previously, are summarized in Table 11. Summaries of the research results of each study can be found in Appendix G.

Table 11. Research Results from Reviewed Research

Type of Accommodation	Research Results	Number of Studies
Computer Administration (N = 9)	Positive effect on scores	4
	No significant effect on scores	3
	Altered item comparability	2
Oral Presentation (N = 10)	Positive effect on scores	6
	No significant effect on scores	1
	Altered item comparability	2
	Did not alter item comparability	1
Extended Time (N = 7)	Positive effect on scores	4
	No significant effect on scores	3
Student-Paced Video (N = 1)	Positive effect on scores	1
Examiner Familiarity (N = 1)	Positive effect on scores	1
Type of Calculator (N = 1)	No significant effect on scores	1
Simplified Language (N = 1)	No significant effect on scores	1
Sign Language (N = 1)	Altered item comparability	1
Meta Analyses of Accommodated Conditions (N = 5)	Positive effect on scores	5
Differential Item Functioning (DIF) Under Accommodated Conditions (N = 6)	Some items exhibited DIF under accommodated conditions	6
Educator Beliefs (N = 4)	Accommodation decisions are based on educator beliefs	4

Read Aloud. The greatest number of studies analyzed oral administration, often referred to as a "read aloud" accommodation. This accommodation was generally found to have positive effects on test scores of students with disabilities. For example, Calhoon et al. (2000) found that students performed better when a teacher read an assessment out loud than when standard paper/pencil administration was used. Helwig et al. (1999) found that students with low math proficiency, regardless of reading ability, scored better with oral presentation of a math test. Only one study did not result in a significant effect on test scores. Two additional studies found that oral administration altered item comparability, affecting the construct the assessment was intended to measure, while one other study did not result in alterations in item comparability.

Computer Administration. The nine studies conducted on the use of computer administration as an accommodation showed varied results. Four of these studies found a positive effect on scores. For example, Brown and Augustine (2001) found that students with reading disabilities performed better using screen reading software than on paper/pencil tests. Burk (1999) also found that students performed better when they used a computer. Three studies resulted in no significant effects on scores. Two studies found that the use of computer administered tests altered item comparability, affecting the construct the assessment was intended to measure.

Extended Time and Multiple Days. The use of extended time and the administration of tests over multiple days had a positive effect on the scores of students with disabilities in four studies. For example, Fuchs et al. (2000a) found that performance on problem-solving curriculum-based measures improved for students with learning disabilities using the extended time accommodation. Similarly, Huesman and Frisbie (2000) found that students with learning disabilities made significantly greater gains on a norm-referenced test with extended time than students without learning disabilities. However, four additional studies did not find an effect of extended time or multiple-day testing on the scores of students with disabilities.

Calculator Use, Encoding, and Examiner Familiarity. Positive effects were found on studies of calculator use, encoding, and examiner familiarity. For example, Fuchs et al. (2000a) found that performance on problem-solving curriculum based measures improved for students with learning disabilities using a calculator and using encoding (writing responses for students). Szarko (2000) found that examiner familiarity had a significant positive effect on the behavior and testing performance of children with autism.

Meta Analyses and Other Studies. All of the meta analyses and studies of multiple accommodations found positive effects of accommodations on test scores of students with disabilities. For example, a meta analysis by Chiu and Pearson (1999) found that the use of accommodations improved test scores of students with disabilities. Elliott, S. et al. (2001) found that performance with test accommodation packages resulted in moderate to large positive effects on test scores of students with disabilities. Similarly, Schulte et al. (2001) found that students with disabilities receiving accommodation packages other than extra time and oral presentation experienced a significant and differential impact of testing accommodations on math scores.

Six studies resulted in Differential Item Functioning (DIF), which occurs when students equated on relevant ability (as defined by test performance), but representing different groups have statistically-defined different probabilities of responding correctly to test items. DIF is investigated by comparing item difficulty. For example, Lewis et al. (1999) found that for both reading and math assessments, only a few items exhibited differential item functioning (DIF) for participants under the reading accommodation conditions; more English language arts items than math items exhibited DIF.

Finally, four studies found that the beliefs of teachers about accommodation use influenced the selection of instructional and assessment accommodations for students with disabilities.

Limitations Cited by Researchers

Limitations of the research were discussed in 21 of the articles reviewed for this analysis. There were three primary limitations identified (see Table 12), including unknown variations among

students included in the study, sample sizes too small to provide adequate statistical support, and nonstandard administration of the accommodations across proctors and schools. Among the wide-ranging unknown variations cited by researchers were type of disability, the withholding of typically-used accommodations, and self-selection biases. Appendix H contains a list of limitations found across the studies.

Table 12. Research Limitations

Limitation	Number of Studies (out of 21 reporting limitations)*
Unknown Variations Between Students	11
Small Sample Size	9
Nonstandard Administration Across Proctors and Schools	6

*Some researchers cited limitations in more than one category.

Recommendations for Future Research

Recommendations for future research were made in 21 of the articles (see Table 13). The recommendations ranged from suggestions by 11 authors to replicate the research for validation and generalization, to investigating associations to specific disabilities, conducting more detailed non-experimental studies to provide richer data, increasing researcher control of the testing process, and studying larger groups of students. These recommendations are described further in Appendix H.

Table 13. Recommendations for Future Research

Recommendations	Number of Studies (out of 21 listing recommendations)*
Replicate Results for Validation and Generalization	11
Investigate Specific Disability Associations	4
Conduct More Detailed Non-experimental Studies to Provide Richer Data	3
Increase Researcher Control of Testing Process	2
Study Larger Sample	1

* Some researchers did not make recommendations; others made more than one recommendation.

Discussion and Implications for Future Research

Several important observations are evident from the analysis of the 46 studies included in this synthesis report. These observations are not conclusive, but can provide direction for future research, policy, and practice. We discuss here some of primary observations from our analysis, with a discussion of implications for the future.

Over half of the studies examined the effects of the use of accommodations on test scores. While this purpose continues to be important, additional studies are needed that investigate the effects of accommodations under much more carefully defined conditions. Specifically, there is a need for clear definition of the constructs tested – not just for the test in general, but for each and every item. There needs to be corroborating information that the intended construct measured does indeed get assessed by each item. At the same time, greater clarity in the accommodations needed by individual students needs to be added – independent ways of measuring whether each student who participates in an accommodation study actually needs the accommodation being studied. Once this clarity is obtained, then better studies of test score validity can be conducted. These should look at both the extent to which the use of an accommodation increases the scores of students who need them, as well as result in better measurement of the students' knowledge and skills – measurement comparable to that obtained for students without disabilities.

Studies are also needed that explore the desirability and perceived usefulness of accommodations by students themselves – the "end users" of assessments. Several researchers cited a lack of information about the individual students who used accommodations as a limitation; in fact, they often expressed frustration about not really knowing whether individual students actually needed the accommodations that were provided to them. Research in which random or inadequate decisions are made about who should use an accommodation, or research in which students are using an accommodation for the first time, may result in limited validity. Researchers also need to consider the implications of multiple accommodation use. Most students use a combination of accommodations during an assessment (Bielinski, Ysseldyke, Bolt, Friedebach & Friedebach, 2001; Brown & Augustine, 2001; Elliott, Bielinski, Thurlow, DeVito, & Hedlund, 1999) and may not do well if only one accommodation is provided for research situations.

Almost half of the studies used state-level criterion-referenced tests or performance assessments, with fewer using norm-referenced or school-designed tests. Because of the importance of the use of criterion-referenced or standards-based tests for accountability purposes (as required by the No Child Left Behind Act of 2001), accommodations research using these tests continues to be the most relevant for states. In addition, the majority of the studies used assessments that addressed the content areas of mathematics and English language arts. Since NCLB requires states to assess students in science by 2007-2008, increased research in science will be important.

Over one third of the studies focused on the accommodation of extended time; in fact the majority of accommodations research over the past several years examined extended time. Extended time is an issue for students who are taking norm-referenced tests, which have traditionally been timed tests; however, since most state tests are criterion-referenced and do not have time limitations, and since this research has fairly consistently concluded that extended time helps students with disabilities, it is time to move on and do less research on extended time and more on other, more controversial and highly used accommodations. For example, oral test administration is a very important and controversial accommodation for students with a variety of disabilities and needs to continue.

Another growing concern is the use of computer-based testing, not just as an accommodation, but for all students. The advent of computer-based testing will bring new challenges for students with disabilities. Research in this area has begun, but will be heightened in intensity and importance as these assessments are developed and used across states (Thompson, Thurlow, Quenemoen, & Lehr, 2002). Research is specifically needed on the use of several features that do not apply to paper/pencil tests, such as familiarity with computer use, screen navigation, screen readers (a variation of oral presentation), and the use of speech recognition software.

There are several considerations for future research in the selection of the participant sample. First is the number of students needed for an adequate sample. The number of participants in the studies examined in this paper ranged from less than one hundred to several thousand. Though the optimal number of subjects varies somewhat with the research design, several researchers cited an inadequate sample size as a limitation of their results. A second consideration is the percent of the sample that consists of students with disabilities. This also varied greatly across studies – from less than 25% of the sample to 100%, and in eight studies the percentage was unknown. It is important to have approximately at least as many students with disabilities as without, especially in studies using research designs that compare the effects of accommodations between the two groups.

In addition to the percent of students with and without disabilities, it is important to have information about the specific disabilities experienced by study participants. Most of the studies examined students with learning or other cognitive deficits. Although it is important to focus research on the largest number of students affected by accommodations use, additional research is needed on accommodations use by students with visual, hearing, and physical disabilities. These students are smaller in number than those with learning disabilities, but often have very complex accommodation needs, including Braille, sign language interpretation, and assistive technology.

Finally, in looking at participant grade levels, it was noted that the majority of studies examined students at the elementary and middle school levels, with very few at the high school level. The

greatest number of studies took place at the fourth grade level. Students at this age may not have command of many accommodations – factors such as inexperience with accommodations and large-scale tests may affect research results.

Over one third of the studies (17) applied non-experimental research designs rather than reflecting one of the group or single subject designs described in Thurlow, et al. (2000). More rigorous research, using designs comparing scores and interactions between the presence and absence of a disability are needed in the future.

Given the increased emphasis placed on scientifically-based findings, the need for more experimental designs is obvious. These designs allow for clearer discrimination of the effects of accommodations, including the isolation of the effect of specific accommodations. Although, experimental designs are best, it is not always an option to employ these designs. The benefits of non-experimental research are: (1) large sample sizes, sometimes even all students in a state, and (2) real-world testing situations (i.e., results should reflect what actually takes place in real-world testing). Within the domain of accommodations research, non-experimental research can play a vital role – addressing the question of comparability in a way not possible with most experimental studies.

The results of the research examined in this paper vary, but show some consistencies that are worth noting. First of all, three accommodations showed a positive effect on student test scores across at least four studies: computer administration, oral presentation, and extended time. However, additional studies on each of these accommodations also found no significant effect on scores or alterations in item comparability. All of the meta analyses of accommodated conditions found a positive effect on scores, and all of the studies examining differential item functioning (DIF) under accommodated conditions found some items that exhibited DIF. Almost all DIF studies expose a few DIF items, whether the comparison is between males and females, different ethnic groups, or accommodated and non-accommodated conditions. So, little DIF usually does not pose a problem, but how many DIF findings are "little"? Clarification of these standards is needed.

Another common finding was that accommodation decisions are based primarily on educator beliefs. The inconsistencies in these decisions have been well-documented (Fuchs, Fuchs, Eaton, Hamlett, & Karns, 2000a; Schulte, Elliott, & Kratochwill, 2000). Identifying specific ways to improve these decisions and to verify them is clearly needed.

The primary recommendation made by researchers for future studies on accommodations is further replication of previously conducted research for increased validation and generalization, with consideration of the other recommendations previously presented in this discussion. It will be important to address the limitations cited repeatedly by researchers: unknown variations between students included in the study, sample sizes too small to provide adequate statistical support,

and nonstandard administration of the accommodations across proctors and schools. Some of these limitations are simply the result of the difficulties inherent in this type of research.

This summary is intended to provide direction to the design of critically needed future research on accommodations use. We are beginning to explore the notion that tests can be designed from the beginning to be better for everyone. Future research should also explore the effects of assessment design and standardization to see whether incorporating new item designs and incorporating more flexible testing conditions reduces the need for accommodations while facilitating measurement of the critical constructs for students with disabilities. It is possible that through implementation of the principles of universal design (Thompson, Johnstone, & Thurlow, 2002), the need for accommodations will decrease, and the measurement of what students know and can perform will improve for *all* students.

References

Barton, K. E., & Huynh, H. (2000). *Patterns of errors made on a reading test with oral reading administration.* Paper presented at the annual conference of the National Council on Measurement in Education, New Orleans, LA.

Bielinski, J., Thurlow, M., Ysseldyke, J., Freidebach, J., & Freidebach, M. (2001). *Read-aloud accommodations: Effects on multiple-choice reading and math items* (Technical Report 31). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Bielinski, J., Ysseldyke, J., Bolt, S., Freidebach, M., & Freidebach, J. (2001). *Prevalence of accommodations for students with disabilities participating in a statewide testing program. Assessment for Effective Intervention 26 (2), 21-28.*

Bourke, A. B., Strehorn, K. C., & Silver, P. (2000). Faculty members' provision of instructional accommodations to students with LD. *Journal of Learning Disabilities, 33(1),* 26-32.

Brown, P. J., & Augustine, A. (2001). *Screen reading software as an assessment accommodation: Implications for instruction and student performance.* Paper presented at the annual meeting of the American Education Research Association, Seattle, WA.

Burk, M. (1999). *Computerized test accommodations.* Washington, D.C.: A.U. Software, Incorporated.

Calhoon, M. B., Fuchs, L. S., & Hamlett, C. L. (2000). Effects of computer-based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly, 23,* 271-282.

Chiu, C. W. T., & Pearson, P. D. (1999). *Synthesizing the effects of test accommodations for special education and limited English proficiency students.* Paper presented at the National Conference on Large Scale Assessment, Snowbird, UT (ERIC Document Reproduction Service No. ED 433 362).

DiCerbo, K. E., Stanley, E., Roberts, M., & Blanchard, J. (2001). *Attention and standardized reading test performance: Implications for accommodation.* Paper session presented at the annual meeting of the National Association of School Psychologists, Washington, DC.

Elliott, J., Bielinski, J., Thurlow, M., DeVito, P., & Hedlund, E. (1999). *Accommodations and the performance of all students on Rhode Island's performance assessment* (Rhode Island Assessment Report 1). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Elliott, S., Kratochwill, T., & McKevitt, B. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology, 39*, 3-24.

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. M. (2000a). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review, 29*, 65-85.

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., Binkley, E., & Crouch, R. (2000b). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children, 67*, 67-81.

Haaf, R., Duncan, B., Skarakis-Doyle, E., Carew, M., & Kapitan, P. (1999). Computer-based language assessment software: The effects of presentation and response format. *Language, Speech, and Hearing Services in Schools, 30*, 68-74.

Hanson, K., Brown, B., Levine, R., & Garcia, T. (2001). Should standard calculators be provided in testing situations? An investigation of performance and preference differences. *Applied Measurement in Education, 14 (1),* 59-72.

Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *The Journal of Educational Research, 93*, 113-125.

Helwig, R., Stieber, S., Tindal, G., Hollenbeck, K., Heath, B., & Almond, P. (2000). *A comparison of factor analyses of handwritten and word-processed writing of middle school students* Eugene: RCTP.

Hollenbeck, K., Rozek-Tedesco, M., Tindal, G., & Glasgrow, A. (2000). An exploratory study of student-paced versus teacher-paced accommodations for large-scale math tests. *Journal of Special Education Technology, 15(2),* 27-36.

Hollenbeck, K., Tindal, G., Harniss, M., & Almond, P. (1999a). *The effect of using computers as an accommodation in a statewide writing test*. Eugene, OR: University of Oregon, BRT.

Hollenbeck, K., Tindal, G., Stieber, S., & Harniss, M. (1999b). *Handwritten versus word-processed statewide compositions: Do judges rate them differently?* Eugene, OR: University of Oregon, RCTP.

Huesman, R., & Frisbie, D. (2000). *The validity of ITBS reading comprehension test scores for learning disabled and non learning disabled students under extended-time conditions*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Johnson, E. S. (2000). The effects of accommodations on performance assessments. *Remedial and Special Education, 21*, 261-267.

Johnson, E., Kimball, K., & Brown, S.O. (2001). American sign language as an accommodation during standards-based assessments. *Assessments for Effective Intervention, 26(2)*, 39-47.

Koretz, D., & Hamilton, L. (1999). *Assessing students with disabilities in Kentucky: The effects of accommodations, format, and subject. (Rep. No. 498)*. Los Angeles, CA: Center for Research on Standards and Student Testing. (ERIC Document Reproduction Services No. ED 440 148).

Kortez, D., & Hamilton, L. (2001). *The performance of students with disabilities on New York's Revised Regents Comprehensive Examination in English* (Technical Report 540). Los Angeles, CA: Center for the Study of Evaluation.

Kortez, D., & Hamilton, L. (2000). Assessment of students with disabilities in Kentucky: Inclusion, student performance, and validity. *Educational Evaluation and Policy Analysis, 22 (3)*, 255-272.

Kosciolek, S., & Ysseldyke, J. E. (2000). *Effects of a reading accommodation on the validity of a reading test* (Technical Report 28). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Lewis, D., Green, D. R., & Miller, L. (1999). *Using differential item functioning analyses to assess the validity of testing accommodated students with disabilities*. Paper presented at the national conference on large-scale assessment, Snowbird, UT.

Marquart, A. (2000). *The use of extended time as an accommodation on a standardized mathematics test: An investigation of effects on scores and perceived consequences for students of various skill levels.* Paper presented at the annual meeting of the Council of Chief State School Officers, Snowbird, UT.

McKevitt, B. (2000). *The use and effects of testing accommodations on math and science performance assessments*. Paper presented at the annual meeting of the Council of Chief State School Officers, Snowbird, UT.

McKevitt, B., Marquart, A., Mroch, A., Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (2000). *Understanding the effects of testing accommodations: A single case approach.* Paper presented at the annual meeting of the Council of Chief State School Officers, Snowbird, UT.

McKevitt, B., Marquart, A., Mroch, A., Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (1999). *Test accommodations for students with disabilities: An empirical analysis.* Poster presented at the Annual Meeting of the American Psychological Association, Boston, MA.

Medina, J. (2000). Classroom testing accommodations for postsecondary students with learning disabilities: The empirical gap. *Dissertation Abstracts International*, 60, 2372.

Meloy, L. L., Deville, C., & Frisbie, D. (2000). *The effect of a reading accommodation on standardized test scores of learning disabled and non learning disabled students.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA (ERIC Document Reproduction Service No. 441 008).

Pomplun, M., & Omar, M. H. (2000). Score comparability of a state mathematics assessment across students with and without reading accommodations. *Journal of Applied Psychology, 85*, 21-29.

Russell, M., & Plati, T. (2000). Effects of computer versus paper administration of a state-mandated writing assessment. TCRecord.org. Retrieved January 23, 2001, from the World Wide Web: http://www.tcrecord.org/PrintContent.asp?ContentID=10709.

Russell, M. (1999). Testing writing on computers: A follow-up study comparing performance on computer and on paper. *Educational Policy Analysis Archives, 7*.

Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (2000). Educators' perceptions and documentation of test accommodations for students with disabilities. *Special Services in the Schools, 16*, 35-56.

Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (2001). Effects of testing accommodations on standardized mathematics test scores: An experimental analysis of the performances of students with and without disabilities. *School Psychology Review, 30(4)*, 527-547.

Simpson, R. L., Griswold, D. E., & Myles, B. S. (1999). Educators' assessment accommodation preferences for students with autism. *Focus on Autism and Other Developmental Disabilities*, 14, 212-219.

Szarko, J. (2000). Familiar versus unfamiliar examiners: The effects on testing performance and behaviors of children with autism and related developmental disabilities. *Dissertation Abstracts International*, 61(4-B).

Thompson, S. J., & Thurlow, M. L. (1999). *1999 state special education outcomes: A report on state activities at the end of the century.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S. J., & Thurlow, M. L. (2001). *2001 State special education outcomes: A report on state activities at the beginning of a new decade.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thompson, S. J., Thurlow, M. L., Quenemoen, R.F., & Lehr, C. (2002). *Access to computer-based testing for students with disabilities.* (Synthesis Report 45). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M., & Bolt, S. (2001). *Empirical support for accommodations most often allowed in state policy* (Synthesis Report 41).Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., Lazarus, S., Thompson, S., & Robey, J. (2002). *2001 state policies on assessment participation and accommodations* (Synthesis Report 46). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Thurlow, M. L., McGrew, K. S., Tindal, G., Thompson, S. L., Ysseldyke, J. E., & Elliott, J. L. (2000). *Assessment accommodations research: Considerations for design and analysis* (Technical Report 26). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Tindal, G., Anderson, L., Helwig, R., Miller, S., & Glasgow, A. (2000). *Accommodating students with learning disabilities on math tests using language simplification.* Eugene, OR: University of Oregon, RCTP.

Vogel, S., Leyser, Y., Wyland, S., & Brulle, A. (1999). Students with learning disabilities in higher education: Faculty attitude and practices. *Learning Disabilities Research & Practice, 14(30),* 173-186.

Walz, L., Albus, D., Thompson, S., & Thurlow, M. (2000). *Effect of a multiple day test accommodation on the performance of special education students* (Minnesota Report No. 34). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Weston, T. (1999). *The validity of oral presentation in testing.* Montreal, Canada: American Educational Research Association.

Zurcher, R.(1999). The effects of testing accommodations on the admissions test scores of students with learning disabilities. *Dissertation Abstracts International*, 90/09, 3324.

Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education, 13(1),* 99-117.

Appendix A

Summary of Research Purpose

Determine the effect of the use of accommodations on test scores of students with disabilities	
Brown, P. J. & Augustine, A. (2001). *Screen reading software as an assessment accommodation: Implications for instruction and student performance.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.	Determine whether assessment items administered using screen reading software are better at measuring student learning than assessment items administered in a paper and pencil format.
Burk, M. (1999). *Computerized test accommodations* Washington, DC: A.U. Software, Incorporated.	Assess the overall efficiency and desirability of a computerized testing environment for persons with disabilities, and systematically examine the effects of certain modifications on test performance.
Calhoon, M. B., Fuchs, L. S., & Hamlett, C. L. (2000). Effects of computer-based test accommodations on mathematics performance assessments for secondary students with learning disabilities. *Learning Disability Quarterly, 23*, 271-282.	Compare the effects of several test accommodations (standard administration, teacher-read text, computer-read, and computer read with video) on math performance assessment scores for secondary students with learning disabilities.
Chiu, C. W. T., & Pearson, P. D. (1999). *Synthesizing the effects of test accommodations for special education and limited English proficiency students.* Paper presented at the National Conference on Large Scale Assessment, Snowbird, UT (ERIC Document Reproduction Service No. ED 433 362).	Determine what constitutes an effective accommodation, the extent to which an accommodation can improve the scores of target students, and whether accommodations should be provided to only target students or to all students.
DiCerbo, K. E., Stanley, E., Roberts, M., & Blanchard, J. (2001). *Attention and standardized reading test performance: Implications for accommodation.* Paper session presented at the annual meeting of the National Association of School Psychologists, Washington, DC.	Determine whether the number of days over which a test was administered (one or three) affects the test performance of subjects with differing reading abilities on a test of comprehension.
Elliott, J., Bielinski, J., Thurlow, M., DeVito, P., Hedlund, E. (1999). *Accommodations and the performance of all students on Rhode Island's performance assessment (Rhode Island Assessment Report 1).* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.	Address the effects of accommodations and the inclusion of all students in Rhode Island's 1996 State Assessment Program. Examine the number of students using specific accommodations and some of the technical characteristics of tests when administered to students with disabilities.
Elliott, S., Kratochwill, T., & McKevitt, B. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology, 39*, 3-24.	Document the testing accommodations educators use when assessing students with performance assessment tasks and investigate the effects these accommodations have on an individual student's test performance.
Hanson, K., Brown, B., Levine, R., & Garcia, T. (2001). Should standard calculators be provided in testing situations? An investigation of performance and preference differences. *Applied Measurement in Education, 14 (1)*, 59-72.	Examine the extent to which test performance of students from diverse backgrounds varies depending on whether students use their own or standard calculators.
Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *The Journal of Educational Research, 93*, 113-125.	Examine the effect of providing middle school students with a video accommodation for a standardized mathematics test.

Determine the effect of the use of accommodations on test scores of students with disabilities	
Helwig, R., Stieber, S., Tindal, G., Hollenbeck, K., Heath, B., & Almond, P. (2000). *A comparison of factor analyses of handwritten and word-processed writing of middle school students*. Eugene, OR: University of Oregon, RCTP.	Compare traditional pencil and paper writing with writing done with the aid of a word processor in order to improve decisions about the appropriateness of word processors as an accommodation, modification, or student-elected option on writing assessments.
Hollenbeck, K., Rozek-Tedesco, M., Tindal, G., & Glasgrow, A. (2000). An exploratory study of student-paced versus teacher-paced accommodations for large-scale math tests. *Journal of Special Education, 15(2),* 27-36.	Determine whether a teacher-paced video accommodation or a student-paced computer accommodation provided differential access for students with disabilities.
Hollenbeck, K., Tindal, G., Harniss, M., & Almond, P. (1999a). The effect of using computers as an accommodation in a statewide writing test Eugene, OR: University of Oregon, BRT.	Investigate the best way for using computers as an accommodation for statewide writing assessments and the effect of the computer accommodation on students with disabilities.
Huesman, R., & Frisbie, D. (2000). *The validity of ITBS reading comprehension test scores for learning disabled and non-learning disabled students under extended-time conditions*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.	Uncover the impact of extended-time limits on test performance and score comparability for the Iowa Test of Basic Skills (ITBS) Reading Comprehension scores for learning disabled (LD) and non-learning disabled (NLD) students.
Johnson, E. S. (2000). The effects of accommodations on performance assessments. *Remedial and Special Education, 21*, 261-267.	Determine whether reading the mathematics items on a test significantly affects the scores of students with and without learning disabilities.
McKevitt, B., Marquart, A., Mroch, A., Gilbertson Schulte, A., Elliott, S., & Kratochwill, T. (2000). *The use and effects of testing accommodations on math and science performance assessments*. Paper presented at the annual meeting of the National Association of School Psychologists, New Orleans, LA.	Describe research that contributes to existing knowledge about the use and effects of testing accommodations for students with disabilities.
McKevitt, B., Marquart, A., Mroch, A., Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (1999). *Test accommodations for students with disabilities: An empirical analysis*. Poster presented at the annual meeting of the American Psychological Association, Boston, MA.	Describe research that enhances knowledge about the use and effects of testing accommodations.
McKevitt, B., Marquart, A., M. A., Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (2000). *Understanding the effects of testing accommodations: A single case approach*. Paper presented at the annual meeting of the CCSSO Large-Scale Assessment Conferences, Snowbird, UT.	Describe single-case research that adds to the knowledge about the use and effects of testing accommodations for students with disabilities.
Medina, J. (2000). Classroom testing accommodations for postsecondary students with learning disabilities: The empirical gap. *Dissertation Abstracts International*, 60(7-A).	Evaluate the effect of extended testing time on classroom tests for postsecondary students with and without disabilities.
Meloy, L. L., Deville, C., & Frisbie, D. (2000). *The effect of a reading accommodation on standardized test scores of learning disabled and non-learning disabled students*. Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA (ERIC Document Reproduction Service No. 441 008).	Examine the effect of the Read Aloud accommodation on the performance of middle school students with learning disabilities in the area of reading (LD-R) and students without learning disabilities (non LD) using tests from the Iowa Tests of Basic Skills achievement battery.
Russell, M. (1999). Testing writing on computers: A follow-up study comparing performance on computer and on paper. *Educational Policy Analysis Archives*, 7.	Examine the effect of taking open-ended tests on computers and on paper for students with different levels of computer skill.

Determine the effect of the use of accommodations on test scores of students with disabilities	
Russell, M. & Plati, T. (2000). Effects of computer versus paper administration of a state-mandated writing assessment. TCRecord.org. Retrieved January 23, 2001, from the World Wide Web: http://www.tcrecord.org/PrintContent.asp?ContentID	Build on research that investigates the effect that computer versus paper-and-pencil administration has on student performance on open-ended items requiring written responses.
Schulte, A.G., Elliott, S., & Kratochwill. T. (2001). Effects of testing accommodations on standardized mathematics test scores: An experimental analysis of the performance of students with and without disabilities. *School Psychology Review*, 30(4), 527-547.	Examine the use and effect of testing accommodations on the test scores of students with and without disabilities on corresponding forms of a mathematics test used in many statewide assessment programs.
Szarko, J. (2000). Familiar versus unfamiliar examiners: The effects on testing performance and behaviors of children with autism and related developmental disabilities. *Dissertation Abstracts International, 61(4-B).*	Determine the effects of examiner familiarity on the behavior and testing performance of children with autism.
Walz, L., Albus, D., Thompson, S., & Thurlow, M. (2000). *Effect of a multiple day test accommodation on the performance of special education students (Minnesota Report 34).* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.	Determine the effect of allowing students to take a reading test over multiple days opposed to taking a reading test on one day.
Investigate the effects of accommodations on test score validity	
Bielinski, J., Thurlow, M., Ysseldyke, J., Freidebach, J., & Freidebach, M. (2001). *Read-aloud accommodation: Effects on multiple-choice reading & math items* (Technical Report No. 31). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.	Examine the effects of the read aloud accommodation on reading and math test score validity.
Haaf, R., Duncan, B., Skarakis-Doyle, E., Carew, M., & Kapitan, P. (1999). Computer-based language assessment software: The effects of presentation and response format. *Language, Speech, and Hearing Services in Schools, 30*, 68-74.	Illustrate the equivalency of two computer-based modifications of a standardized receptive language test with a sample of typically developing children.
Johnson, E., Kimball, K., & Brown, S.O. (2001). American sign language as an accommodation during standards-based assessments. *Assessment For Effective Intervention, 26(2),* 39-47.	Investigate the impact of American Sign Language (ASL) as an accommodation on the validity of standards-based assessments.
Kosciolek, S., & Ysseldyke, J. E. (2000). *Effects of a reading accommodation on the validity of a reading test* (Technical Report 28). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.	Identify whether the read aloud accommodation can be offered to students with disabilities without endangering the validity of norm-referenced, standardized high-stakes tests.
Lewis, D., Green, D. R., & Miller, L. (1999). *Using differential item functioning analyses to assess the validity of testing accommodated students with disabilities.* Paper presented at the annual CCSSO Large-Scale Assessment Conference, Snowbird, UT.	Determine whether accommodations alter what is measured by an assessment.
Pomplun, M. & Omar, M. H. (2000). Score comparability of a state mathematics assessment across students with and without reading accommodations. *Journal of Applied Psychology, 85*, 21-29.	Investigate the meaning of math test scores for students with learning disabilities who were read an assessment in comparison with general education students with learning disabilities for whom the assessment was not read.
Tindal, G., Anderson, L., Helwig, R., Miller, S., & Glasgow, A. (2000). *Accommodating students with learning disabilities on math tests using language simplification.* Eugene, OR: University of Oregon, RCTP.	Investigate whether simplifying language changed the construct measured by a math test.

Investigate the effects of accommodations on test score validity	
Weston, T. (1999). *The validity of oral presentation in testing.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.	Establish whether construct irrelevant variance is reduced in test scores of learning disabled students through the use of the testing accommodation of oral presentation.
Zurcher, R. (1999).The effects of testing accommodations on the admissions test scores of students with learning disabilities. *Dissertation Abstracts International*, 90/09, 3324.	Investigate the comparability and criterion validity of the scores from admissions tests after accommodations are provided to individuals with learning disabilities.
Zuriff, G.E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education, 13* (1), 99-117.	Examine the validity of scores on an examination taken by a learning-disabled student under conditions of extended time.
Study institutional factors, teacher judgment, or student desirability of accommodation use	
Bourke, A.B., Strehorn, K.C., & Silver, P. (2000). Faculty members' provision of instructional accommodations to students with LD. *Journal of Learning Disabilities, 33(1),* 26-32.	Investigate possible institutional factors that either facilitate or hinder the accommodation process.
Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. M. (2000a). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review, 29*, 65-85.	Compare the effects of accommodations for students with and without learning disabilities on math curriculum based measures using a variety of accommodations. Examine the usefulness of teacher judgments in specifying accommodations. Investigate whether standardized data-based rules for formulating individual accommodation decisions supplement teacher judgments about accommodations.
Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., Binkley, E., & Crouch, R. (2000b). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children, 67*, 67-81.	Compare the effects of three reading test accommodations for students with and without learning disabilities, examine the usefulness of teacher judgments in specifying accommodations, and investigate whether standardized data-based rules for making individual accommodation decisions might be useful to supplement teacher judgments about accommodations.
Marquart, A. (2000). *The use of extended time as an accommodation on a standardized mathematics test: An investigation of effects on scores and perceived consequences for students of various skill levels.* Paper presented at the annual CCSSO Large-Scale Assessment Conference, Snowbird, UT.	Examine the effects of extended time on the performance of students with disabilities, students academically at-risk, and students functioning at or below grade-level. Analyze student reactions to the use of extended time.
Schulte, A. G., Elliott, S. N., & Kratochwill, T. R. (2000). Educators' perceptions and documentation of test accommodations for students with disabilities. *Special Services in the Schools, 16*, 35-56.	Investigate the differences in the accommodations that educators might offer to students with various disabilities.
Simpson, R. L., Griswold, D. E., & Myles, B. S. (1999). Educators' assessment accommodation preferences for students with autism. *Focus on Autism and Other Developmental Disabilities, 14*, 212-219.	Investigate the views of special educators about the participation of students with autism on districtwide assessments, and accommodations they judge most appropriate.
Vogel, S., Leyser, Y., Wyland, S., & Brulle, A. (1999). Students with learning disabilities in higher education: Faculty attitude and practices. *Learning Disabilities Research & Practice, 14(30),* 173-186.	Examine faculty attitudes and practices toward providing teaching and examination accommodations for students with learning disabilities in higher education.

Examine patterns of errors across items or tests	
Barton, K. E., & Huynh, H. (2000). *Patterns of errors made on a reading test with oral reading administration.* Paper presented at the annual conference of the National Council on Measurement in Education, New Orleans, LA.	Investigate the patterns of errors made by accommodated students with different types of disabilities on a large-scale reading test.
Hollenbeck, K., Tindal, G., Stieber, S., & Harniss, M. (1999b). *Handwritten versus word-processed statewide compositions: Do judges rate them differently?* Eugene, OR: University of Oregon, BRT.	Investigate whether word-processed essays are rated differently than handwritten essays.
Koretz, D., & Hamilton, L. (1999). *Assessing students with disabilities in Kentucky: The effects of accommodations, format, and subject (Technical Report No. 498).* Los Angeles, CA: Center for Research on Standards and Student Testing. (ERIC Document Reproduction Service No. ED 440 148).	Explore Kentucky's efforts to include students with disabilities in its statewide assessment with a specific focus on the state's measurement techniques.
Koretz, D., & Hamilton, L. (2000). Assessment of students with disabilities in Kentucky: Inclusion, student performance, and validity. *Educational Evaluation and Policy Analysis, 22 (3),* 255-272.	Explore Kentucky's efforts to include students with disabilities in its statewide assessment specifically, the level of inclusion achieved, the types of assessment accommodations offered, the performance of students with disabilities, and the relationships between performance and the use of accommodations on multiple-choice and open-response formats.
Koretz, D., & Hamilton, L. (2001). *The performance of students with disabilities on New York's Revised Regents Comprehensive Examination in English (Technical Report 540).* Los Angeles, CA: Center for the Study of Evaluation.	Analyze the performance of students with disabilities on the first field test of New York State's revised Regents Comprehensive Examination in English.

Appendix B

Summary of Type of Assessment

Authors (year)	Norm-Referenced and Other Standardized Tests	State Criterion Referenced Tests or Performance Assessments	School or District-designed Tests	Other (name)
Barton & Huynh (2000)		South Carolina Basic Skills Assessment Program (BSAP) High School Examination Test of Reading, the Oral Accommodation (OA) form		
Bielinski et al. (2001)		The Missouri Assessment Program (MAP)		
Bourke et al. (2000)				Survey
Brown & Augustine (2001)	Delaware and Pennsylvania NAEP (National Assessment of Educational Progress)			
Burk (1999)	Abbreviated forms of the Test of General Educational Development (GED) practice test	Abbreviated version of the State of Maryland Functional Math Test	Transition planning program post-test	
Calhoon et al. (2000)			Mathematics Performance Assessment	
Chiu & Pearson (1999)				Multiple assessments*
DiCerbo et al. (2001)	Stanford Achievement Test, Ninth Edition (SAT-9)			
Elliott, J. et al. (1999)		Rhode Island Performance Assessments		
Elliott, S. et al. (2001)		Performance Assessment Instruments Developed from the Wisconsin Student Assessment System Project		
Fuchs et al. (2000a)	Iowa Tests of Basic Skills (ITBS), Stanford Achievement Test (SAT)		Computational Curriculum Based Measures	
Fuchs et al. (2000b)	Iowa Tests of Basic Skills (ITBS)		Reading Passages from Monitoring Basic Skills Progress	

Authors (year)	Norm-Referenced and Other Standardized Tests	State Criterion Referenced Tests or Performance Assessments	School or District-designed Tests	Other (name)
Haaf et al. (1999)	Peabody Picture Vocabulary Test-Revised (PPVT-R)			
Hanson et al. (2001)	National Assessment of Educational Progress (NAEP) problem sets, Timed math calculation tests			
Helwig et al. (1999)	Standardized Mathematics Achievement Test			
Helwig et al. (2000)		Compositions were evaluated using Oregon's 6-Trait Analytic Scoring System		
Hollenbeck et al. (2000)		Multiple-choice test based on the item pool from an actual statewide test		
Hollenbeck et al. (1999a)		Statewide Test of Writing Proficiency		
Hollenbeck et al. (1999b)		Writing portion of the Oregon Statewide Assessment		
Huesman & Frisbie (2000)	Iowa Test of Basic Skills (ITBS)			
Johnson (2000)		Washington Assessment of Student Learning (WASL)		
Johnson et al. (2001)		Washington Assessment of Student Learning (WASL)		
Koretz & Hamilton (1999)		Kentucky Instructional Results Information System (KIRIS)		
Koretz & Hamilton (2000)		Kentucky Instructional Results Information System (KIRIS)		
Koretz & Hamilton (2001)		The New York Regents Comprehensive Examination in English		
Kosciolek & Ysseldyke (2000)	California Achievement Tests Comprehension Survey			
Lewis et al. (1999)		Indiana Statewide Testing for Educational Progress Graduation Qualifying Exam (ISTEP+GQE)		

Authors (year)	Norm-Referenced and Other Standardized Tests	State Criterion Referenced Tests or Performance Assessments	School or District-designed Tests	Other (name)
Marquart (2000)	Alternate Forms of Standardized Math Tests Developed from the TerraNova Level 18 Mathematics Test			
McKevitt (2000)		Performance assessments from the Wisconsin Student Assessment System 1993-1995		
McKevitt et al. (1999)		Performance assessments from the Wisconsin Student Assessment System Project 1993-1995		
McKevitt et al. (2000)		Performance assessments from the Wisconsin Student Assessment System Project 1993-1995		
Medina (2000)	Scholastic Abilities Test for Adults			
Meloy et al. (2000)	Iowa Tests of Basic Skills (ITBS)			
Pomplun & Omar (2000)		Kansas Assessment Program		
Russell (1999)	Stanford Achievement Test, Ninth Edition (SAT-9)			
Russell & Plati (2000)		Massachusetts Comprehensive Assessment System (MCAS)		
Schulte et al. (2000)				Assessment Accommodation Checklist
Schulte et al. (2001)	TerraNova Multiple Assessment Battery			
Simpson et al. (1999)				Survey
Szarko (2000)	Psychoeducational Profile-Revised (PEP-R)			
Tindal et al. (2000)			Two forms of a math test matched for difficulty and problem type	

Authors (year)	Norm-Referenced and Other Standardized Tests	State Criterion Referenced Tests or Performance Assessments	School or District-designed Tests	Other (name)
Vogel et al. (1999)				Survey
Walz et al. (2000)		Measures Based on the Minnesota Basic Standards Tests (BSTS)		
Weston (1999)			A math test	
Zurcher (1999)	Miller Analogies Test			
Zuriff (2000)				Multiple assessments*
Total	17	21	6	6

* Indicates a review or meta analysis of multiple studies.

Appendix C

Summary of Content Area Assessed

Authors (year)	Reading/ Language Arts	Writing	Mathematics	Science	Social Studies	No Specific Content Area	Total
Barton & Huynh (2000)	X						1
Bielinski et al. (2001)	X		X				2
Bourke et al. (2000)						X	1
Brown & Augustine (2001)				X	X		2
Burk (1999)						X	1
Calhoon et al. (2000)			X				1
Chiu & Pearson (1999)						X	1
DiCerbo et al. (2001)	X						1
Elliott, J. et al. (1999)	X	X	X				3
Elliott, S. et al. (2001)			X	X			2
Fuchs et al. (2000a)			X				1
Fuchs et al. (2000b)	X						1
Haaf et al. (1999)	X						1
Hanson et al. (2001)			X				1
Helwig et al. (1999)			X				1
Helwig et al. (2000)		X					1
Hollenbeck et al. (2000)			X				1
Hollenbeck et al. (1999a)		X					1
Hollenbeck et al. (1999b)		X					1
Huesman & Frisbie (2000)	X						1
Johnson (2000)		X					1
Johnson et al. (2001)	X		X				2
Koretz & Hamilton (1999)	X		X	X	X		4
Koretz & Hamilton (2000)	X		X	X	X		4
Koretz & Hamilton (2001)	X	X					2

Authors (year)	Reading/ Language Arts	Writing	Mathematics	Science	Social Studies	No Specific Content Area	Total
Kosciolek & Ysseldyke (2000)	X						1
Lewis et al. (1999)	X		X				2
Marquart (2000)			X				1
McKevitt (2000)			X	X			2
McKevitt et al. (1999)			X	X			2
McKevitt et al. (2000)			X	X			2
Medina (2000)						X	1
Meloy et al. (2000)	X		X	X			3
Pomplun & Omar (2000)			X				1
Russell (1999)	X		X	X			3
Russell & Plati (2000)		X					1
Schulte et al. (2000)			X				1
Schulte et al. (2001)			X				1
Simpson et al. (1999)						X	1
Szarko (2000)						X	1
Tindal et al. (2000)			X				1
Vogel et al. (1999)						X	1
Walz et al. (2000)	X						1
Weston (1999)			X				1
Zurcher (1999)						X	1
Zuriff (2000)						X	1
Total	16	7	23	9	3	9	

Appendix D

Summary of Types of Accommodations

	Presentation				Response			Setting	Timing/Scheduling				
Author (year)	Oral Present-ation	Computer Admin.	Simplified Language	Large Print	Dictated Response	Word Processor	Calculator	Separate Setting/ Small Group	Extended Time	Multiple Day	Frequent Breaks	Multiple**	Other
Barton & Huynh (2000)*	X												
Bielinski et al. (2001)*	X							X	X			X	
Bourke et al. (2000)*													X
Brown & Augustine (2001)*	X	X											
Burk (1999)*	X	X		X								X	X
Calhoon et al. (2000)*	X	X											
Chiu & Pearson (1999)*		X						X	X				X
DiCerbo et al. (2001)										X			
Elliott, J. et al. (1999)*	X				X			X	X		X	X	X
Elliott, S. et al. (2001)*									X			X	X
Fuchs et al. (2000a)*	X				X		X		X			X	
Fuchs et al. (2000b)*	X			X					X				X
Haaf et al. (1999)*		X											
Hanson et al. (2001)*							X						

	Presentation				Response			Setting	Timing/Scheduling				
Author (year)	Oral Present-ation	Computer Admin.	Simplified Language	Large Print	Dictated Response	Word Processor	Calculator	Separate Setting/ Small Group	Extended Time	Multiple Day	Frequent Breaks	Multiple**	Other
Helwig et al. (1999)*	X	X											
Helwig et al. (2000)						X							
Hollenbeck et al. (2000)*	X	X											
Hollenbeck et al. (1999a)*						X							X
Hollenbeck et al. (1999b)						X							
Huesman & Frisbie (2000)*									X				X
Johnson (2000)	X												
Johnson et al. (2001)*													X
Koretz & Hamilton (1999)*	X		X		X							X	X
Koretz & Hamilton (2000)*	X		X		X							X	X
Koretz & Hamilton (2001)*	X				X			X	X			X	X
Kosciolek & Ysseldyke (2000)*	X												
Lewis et al. (1999)*	X								X			X	
Marquart (2000)									X				

	Presentation				Response			Setting	Timing/Scheduling				
Author (year)	Oral Present-ation	Computer Admin.	Simplified Language	Large Print	Dictated Response	Word Processor	Calculator	Separate Setting/ Small Group	Extended Time	Multiple Day	Frequent Breaks	Multiple**	Other
McKevitt (2000)*	X		X						X			X	X
McKevitt et al. (1999)*									X			X	X
McKevitt et al. (2000)*	X		X						X			X	X
Medina (2000)									X				
Meloy et al. (2000)*	X								X				
Pomplun & Omar (2000)	X												
Russell (1999)*		X											
Russell & Plati (2000)						X							
Schulte et al. (2000)*													X
Schulte et al. (2001)*	X		X		X			X	X		X	X	X
Simpson et al. (1999)*													X
Szarko (2000)													
Tindal et al. (2000)			X										
Vogel et al. (1999)*													X

	Presentation				Response			Setting	Timing/Scheduling				
Author (year)	Oral Present-ation	Computer Admin.	Simplified Language	Large Print	Dictated Response	Word Processor	Calculator	Separate Setting/ Small Group	Extended Time	Multiple Day	Frequent Breaks	Multiple**	Other
Walz et al. (2000)										X			
Weston (1999)	X												
Zurcher (1999)*												X	
Zuriff (2000)*									X				
Total	22	8	6	2	6	4	2	5	17	2	2	14	19

* See Specification Table
** Accommodations were provided in packages, combinations, groups, or pairs.

Specifications for Appendix D

Authors (year)	Specification(s)
Barton & Huynh (2000)	All students were read the test by a reader, an audiotape operated by either a test administrator or by the student. A videotape of a signed administration was provided for students with hearing impairments.
Bielinski et al. (2001)	A group of students with disabilities received the read aloud accommodation. Many of these students received this accommodation with extended time and in a small group.
Bourke et al. (2000)	Teachers responded to survey items about the helpfulness of and need for instructional accommodations.
Brown & Augustine (2001)	One version of the test was administered via a computer utilizing screen reading software.
Burk (1999)	All participants received computer administration, and some received large print, extra spacing, sound (including a recorded voice reading the test), or a combination of these accommodations.
Calhoon et al. (2000)	Each student completed the tests under each of the subsequent four conditions: standard administration, teacher read aloud, computer read aloud, and computer read aloud with video.
Chiu & Pearson (1999)	Multiple accommodations were examined in this meta analysis, including: extended time/unlimited time, presentation format accommodations, assistive devices, response format accommodations, and setting changes.
Elliott, J. et al. (1999)	Participants received a variety of accommodations and combinations of accommodations. These included extended time, oral presentation, repeated directions, separate setting, oral response, more frequent breaks, and a combination of these accommodations.
Elliott, S. et al. (2001)	Students received either a standard package of accommodations (extra time, support with understanding directions and reading words, verbal encouragement), or accommodations recommended by their teachers or implied by their individualized education programs.
Fuchs et al. (2000a)	Accommodations provided included extended time, calculators, read aloud, and encoding (writing responses for the student upon request). Some of the conditions included more than one accommodation at a time.
Fuchs et al. (2000b)	Students completed four brief reading assessments under four conditions: (1) standard, (2) extended time, (3) large print, (4) student reads aloud. Then, students took a test using accommodation(s) that were determined by teacher or data-based judgment to meet their individual needs.
Haaf et al. (1999)	Computer presentation trackball and computer presentation-automated scanning were studied.
Hanson et al. (2001)	Students completed test under two conditions: (1) using a standard calculator, and (2) using his/her own calculator.
Helwig et al. (1999)	Half of the items on the test were presented on a screen and read aloud over monitor speakers.
Hollenbeck et al. (2000)	The test was administered under two conditions: (1) a teacher-paced video accommodation, (2) a student-paced computer administration. In both conditions, the items were presented individually and read aloud to students.
Hollenbeck et al. (1999a)	Some students took the writing test using traditional methods, using the computer to complete the entire three-day test, handwriting for two days, using the computer one day, handwriting for two days, or using the computer with spell-checker one day.
Huesman & Frisbie (2000)	In addition to receiving extended time some students were told to work at a normal rate and others were told to take their time and work carefully.
Johnson et al. (2001)	American Sign Language was studied.
Koretz & Hamilton (1999)	Multiple accommodations were studied, including those recorded in the Kentucky Instructional Results Informational System (1997). Accommodations offered included oral presentation, paraphrasing, dictating answers, cueing, technological aid, and interpreter. In many instances students received more than one accommodation.

Authors (year)	Specification(s)
Koretz & Hamilton (2000)	Students received a variety of accommodations that were permitted on the Kentucky Instructional Results Information System (KIRIS) and that were determined to be appropriate based on their individual needs. These included dictation, assistance with spelling, oral presentation, paraphrasing, cueing, technological aids, interpreter, and combinations of these accommodations.
Koretz & Hamilton (2001)	Students with disabilities received a variety of accommodations based on their individual needs. These included extended time, separate location, reading/clarifying directions, having the entire test read aloud, spell/grammar check, and a scribe or tape-recorder.
Kosciolek & Ysseldyke (2000)	Students were provided an audiocassette recording of the items and choices read aloud on one form of the test.
Lewis et al. (1999)	Extra time and read aloud were studied individually and in combination. Students received these accommodations with or without additional accommodations.
McKevitt (2000)	Students with disabilities completed half of the tasks with accommodations recommended by teachers on IEPs or Assessment Accommodation Checklists (AAC). In one condition students without disabilities received three common accommodations (reading and paraphrasing directions, providing verbal encouragement, giving extra time), and in another condition these students received accommodations recommended by their teachers on some tasks.
McKevitt et al. (1999)	Two accommodation conditions were studied: a standard accommodations condition (composed of extra time, support understanding directions, and help reading words) and an individualized accommodation condition (teacher recommended accommodations).
McKevitt et al. (2000)	Students with disabilities received accommodations recommended by teachers for some tasks. In one condition, students without disabilities received three accommodations (reading and paraphrasing directions, providing verbal encouragement, and giving extra time). In another condition students without disabilities received accommodations recommended by their teachers on some tasks.
Meloy et al. (2000)	The read aloud administrations took more time than the standard administrations.
Russell (1999)	Open-ended test items were converted to a computer format and administered in two modes, computer and paper and pencil.
Schulte et al. (2000)	Educators completed the Assessment Accommodation Checklist (AAC) to demonstrate which accommodations they considered to be most useful.
Schulte et al. (2001)	The accommodations most commonly used by participants were extra time, separate setting, read aloud, small group administration, positive praise, frequent breaks, encouragement to remain on task, dictated responses, and paraphrase directions/ questions.
Simpson et al. (1999)	Results of a survey of teachers and related service professionals - identified minimally necessary testing accommodations.
Vogel et al. (1999)	Faculty members at a university responded to survey items about their willingness to provide multiple instructional and examination accommodations and their judgments about the fairness of providing accommodations.
Zurcher (1999)	Students with learning disabilities took a test with and without their usual classroom assessment accommodations and they were matched by race/ethnicity, gender, and academic program to students without learning disabilities, who also took the test with the same accommodations as the person they were matched with.
Zuriff (2000)	Five studies on the effects of extra examination time were analyzed.

Appendix E

Summary of Participants

Authors (year)	Number of Study Participants and Percent with Disabilities	Grade-level of Participants	Types of Disabilities of Students Included in the Sample****
Barton & Huynh (2000)	2,924 (87% with disabilities)	High School	Physical disabilities (hearing, orthopedic, vision, and language impairments), emotional disabilities, learning disabilities, educable mentally challenged students
Bielinski et al. (2001)	4,058 math (45% with disabilities) 3,258 reading (49% with disabilities)	Fourth grade (math) Third grade (reading)	Students with an IEP in reading
Bourke et al. (2000)	162	University faculty members	Not applicable
Brown & Augustine (2001)	206 (96 science and 110 social studies)	Twelfth grade	Type of disability not documented
Burk (1999)	182 (73% with disabilities)	Multiple*	Learning disabilities, mild mental retardation, attention deficit disorder
Calhoon et al. (2000)	81 (100% with disabilities)	Ninth through Twelfth grades	Students with math and reading IEP goals
Chiu & Pearson (1999)	NA**	Elementary, middle school, high school, post-secondary**	Not applicable
DiCerbo et al. (2001)	939 (Study 1), 789 (Study 2)	Third grade	Type of disability not documented
Elliott, J. et al. (1999)	Approximately 11,000	Fourth grade	Type of disability not documented
Elliott, S. et al. (2001)	100 (41% with disabilities)	Fourth grade	Learning disability, emotional disturbance, cognitive disability, speech or language disability, autism, health impairment
Fuchs et al. (2000a)	373 (51% with disabilities)	Fourth and Fifth grades	Learning Disability (LD)
Fuchs et al. (2000b)	365 (50% with disabilities)	Fourth and Fifth grades	Learning Disability (LD)
Haaf et al. (1999)	72 (0 with disabilities)	4 years old through 8 years old	No students with disabilities were included in the sample
Hanson et al. (2001)	50 (# with disabilities unknown)	Recently completed Eighth grade	Type of disability not documented
Helwig et al. (1999)	247 (12% with disabilities)	Sixth grade	Type of disability not documented
Helwig et al. (2000)	117 (9% with disabilities)	Eighth grade	Type of disability not documented

Authors (year)	Number of Study Participants and Percent with Disabilities	Grade-level of Participants	Types of Disabilities of Students Included in the Sample****
Hollenbeck et al. (2000)	50 (50% with disabilities)	Seventh grade	Disabilities in the area of math and reading
Hollenbeck et al. (1999a)	164 (27% with disabilities)	Seventh grade	Type of disability not documented
Hollenbeck et al. (1999b)	80 (9% with disabilities)	Average age = 15.1 years	Type of disability not documented
Huesman & Frisbie (2000)	526 (25% with disabilities)	Sixth grade	Learning disability (LD)
Johnson (2000)	115 (33% with disabilities)	Fourth grade	Receiving special education services in reading
Johnson et al. (2001)	Not Reported (100% with disabilities)	Fourth, Seventh, and Tenth grades	Deaf and hard of hearing students
Koretz & Hamilton (1999)	Not Reported (Nearly half of these students had specific learning disabilities)	Fourth, Fifth, Seventh, Eighth, and Eleventh grades	Autism, deaf/blind, multiple disabilities, emotional/behavioral disabilities, mild mental disabilities, physical disabilities/ orthopedically impaired, other health impaired disabilities, traumatic brain injury, hearing impaired, visual disabilities, communications/speech-language disabilities, functional mental disabilities, specific learning disabilities
Koretz & Hamilton (2000)	20,791 (10% served under IDEA)	Fourth, Fifth, Seventh, Eighth, and Eleventh grades	Learning disabilities, speech or language impairments, mental retardation, emotional disturbance
Koretz & Hamilton (2001)	8,750 (6% with disabilities)	Twelfth grade	Learning disabled, multiple disabilities, hearing impaired, other health impaired, visually impaired, emotionally disturbed, orthopedic impaired, speech impaired, mentally retarded, hard of hearing, autistic, deaf/blind
Kosciolek & Ysseldyke (2000)	32 (47% with disabilities)	Third, Fourth, and Fifth grades	Learning disability, emotional behavioral disorder, speech and language disability
Lewis et al. (1999)	Over 3,000 (100% with disabilities)	Tenth grade	Learning disabilities, mental handicaps

Authors (year)	Number of Study Participants and Percent with Disabilities	Grade-level of Participants	Types of Disabilities of Students Included in the Sample****
Marquart (2000)	96 (24% with disabilities)	Eighth grade	Mild learning disabilities, mild physical disabilities, speech and language disabilities, mild cognitive disabilities, behavioral disabilities, emotional disabilities
McKevitt (2000)	218 (33% with disabilities)	Fourth grade	Learning disability, emotional disturbance, cognitive disability, speech/language, other health impairment, autism
McKevitt et al. (1999)	100 (41% with disabilities)	Fourth grade	Type of disability not provided
McKevitt et al. (2000)	143 (41% with disabilities)	Fourth grade	Type of disability not provided
Medina (2000)	235 (# with disabilities unknown)	College	Information not available
Meloy et al. (2000)	260 (24% with disabilities)	Sixth through Eighth grades	Learning Disabilities in Reading (LD-R)
Pomplun & Omar (2000)	3,109 (52% with disabilities)	Fourth grade	Learning Disabilities (LD)
Russell (1999)	382 (# with disabilities unknown)	Sixth through Eighth grades	No information available
Russell & Plati (2000)	289 (# with disabilities unknown)	Eighth and Tenth grades	No information available
Schulte et al. (2000)	118	Educators	Not applicable
Schulte et al. (2001)	86 (50% with disabilities)	Fourth grade	Mild disabilities, mild cognitive disabilities
Simpson et al. (1999)	133	Educators	Not applicable
Szarko (2000)	26 (100% with disabilities)	48 to 88 months	Autism or atypical pervasive developmental disorders
Tindal et al. (2000)	48 (33% with disabilities)	Seventh grade	Specific learning disabilities
Vogel et al. (1999)	420	University faculty members	Not applicable
Walz et al. (2000)	113 (42% with disabilities)	Seventh and Eighth grades	Students receiving services for academic or behavioral needs

Authors (year)	Number of Study Participants and Percent with Disabilities	Grade-level of Participants	Types of Disabilities of Students Included in the Sample****
Weston (1999)	121 (54% with disabilities)	Fourth grade	Learning Disabilities
Zurcher (1999)	30 (50% with disabilities)	College	Learning Disabilities
Zuriff (2000)	NA***	College, graduate, professional school students	Not applicable

*Participants were from public schools, a private residential school, and an adult learning facility affiliated with a public school. Participants included students with learning disabilities, mild mental retardation, and attention deficit disorder, as well as students and adults without disabilities.
**This article is a meta analysis of thirty studies. Participants were different for each of the thirty studies included in the analysis. Participants of the studies examined in this analysis included elementary, post-secondary, middle school, and high school students.
***This article is a review of five different studies. Participants of the studies examined in this article were college and/or graduate/professional school students. Each study included students with and without learning disabilities.
****Information on disability types listed in this column was only provided if specific disabilities were mentioned in the description of the sample.

Appendix F

Summary of Research Designs

Authors (year)	Group Research Design 1	Group Research Design 2	Group Research Design 3	Group Research Design 4	Single Subject Research Designs	Non-Experimental/ Other
Barton & Huynh (2000)						X
Bielinski et al. (2001)						X
Bourke et al. (2000)						X
Brown & Augustine (2001)			X			
Burk (1999)	X					
Calhoon et al. (2000)			X			
Chiu & Pearson (1999)						X
DiCerbo et al. (2001)			X*			
Elliott, J. et al. (1999)						X
Elliott, S. et al. (2001)	X					
Fuchs et al. (2000a)	X**					
Fuchs et al. (2000b)	X					
Haaf et al. (1999)				X***		
Hanson et al. (2001)			X			
Helwig et al. (1999)				X		
Helwig et al. (2000)				X		
Hollenbeck et al. (2000)	X					
Hollenbeck et al. (1999a)		X				
Hollenbeck et al. (1999b)						X
Huesman & Frisbie (2000)						X
Johnson (2000)	X****					
Johnson et al. (2001)						X
Koretz & Hamilton (1999)						X
Koretz & Hamilton (2000)						X
Koretz & Hamilton (2001)						X
Kosciolek & Ysseldyke (2000)	X					

Authors (year)	Group Research Design 1	Group Research Design 2	Group Research Design 3	Group Research Design 4	Single Subject Research Designs	Non-Experimental/ Other
Lewis et al. (1999)						X
Marquart (2000)	X					
McKevitt (2000)					X	
McKevitt et al. (1999)					X	
McKevitt et al. (2000)					X	
Medina (2000)****	X		X			
Meloy & Frisbie (2000)		X				
Pomplun & Omar (2000)						X
Russell (1999)				X		
Russell & Plati (2000)				X		
Schulte et al. (2000)						X
Schulte et al. (2001)			X			
Simpson et al. (1999)						X
Szarko (2000)				X		
Tindal et al. (2000)	X					
Vogel et al. (1999)						X
Walz et al. (2000)			X			
Weston (1999)	X					
Zurcher (1999)	X					
Zuriff (2000)*						X
Total	12	2	7	6	3	17

* Design 3 without counterbalancing
** Design 1 with a single-subject component
*** Design 4 except participants were nondisabled
**** Variation of Design 1 without counterbalancing
***** Authors did not have access to the information necessary to distinguish between Design 1 and 3.

Appendix G

Summary of Research Results

Authors (year)	Computer Use Had Positive Effect On Scores
Burk (1999)	The performance of students with learning disabilities was significantly better on the computerized administration than on the paper-and-pencil version of the tests. No significant differences in performance were found for students with mental retardation or for nondisabled students on the computer administration. For learning disabled students, large print did not appear to have a significant effect on scores, although extra spacing did have a significant positive effect. Students overwhelmingly preferred computerized testing to paper-based testing, regardless of group or disability status, and strongly believed the computer was easier.
Calhoon et al. (2000)	Teacher read, computer read, and computer read with video conditions significantly increased scores obtained on performance assessments completed in the standard administration condition. No significant differences were found among teacher read, computer read, and computer read with video conditions. Effect sizes between the standard condition and each of the accommodated conditions were weak. Students with learning disabilities of varying reading ability levels benefited significantly from the accommodations.
Russell & Plati (2000)	For both grade levels (8 and 10), students who wrote their compositions on the computer produced longer responses and received higher scores for their responses. This effect was significant at both grade levels but largest in grade 8. Since all participants in the sample were found to have a fairly high keyboarding speed, correlations between keyboarding speed and writing composition scores were not strong.
Russell (1999)	A positive group effect was found for the science test. An overall group effect was not found for language arts tests. For students with above average keyboarding speed (0.5 or one-half of a standard deviation above the mean), performance on the computer administered version of the language arts test was significantly higher than performance on "standard administration." In general, performing the math test on the computer had a negative effect, but this effect became less evident as keyboarding speed increased. However, performance on the computer administer language arts test was significantly lower than standard administration for students with poor keyboarding skills.
	Computer Use Had No Significant Effect on Scores
Brown & Augustine (2001)	Format (screen reading versus paper and pencil) did not significantly influence the scores on this assessment when researchers controlled for student reading ability. No significant differences were found between the performance of students completing the pencil and paper format version versus the screen reading format when controlling for reading performance.
Haaf et al. (1999)	No differences in performance were found across conditions. It was concluded that the adapted response formats of the computerized version constitute statistically equivalent forms.
Hollenbeck et al. (1999a)	No significant differences were found between stories written without computers and those composed with a computer. Significant differences were found between ratings for essays of computer-last day group and computer-last day with spell check group. Students with disabilities performed significantly poorer when writing on a computer than when handwriting their stories.

	Computer Use Altered Item Comparability
Helwig et al. (2000)	Single factors emerged when handwritten or word-processed writing samples were examined individually, while two factors emerged when they were viewed together. One factor incorporated all word-processing traits, while the other included all handwritten traits. Correlations revealed a weak relationship across modes, even with common traits. Scores on the handwritten test offered little predictive value of scores received when using a word processor.
Hollenbeck et al. (1999b)	The handwritten compositions, which were transcribed into typed essays before being evaluated by judges, were rated significantly higher statistically than the typed compositions on three of the six traits for the total group. In addition, five of the six mean trait scores were higher for the handwritten essays. Researchers concluded that the two response modes are not comparable and should not be used in the same evaluation system.
	Oral Presentation Had Positive Effect on Scores
Fuchs et al. (2000a)	Students with learning difficulties benefited significantly more than students without LD from the extended time, read aloud, and scribing (writing responses for students) accommodations, and somewhat for calculator use on problem-solving curriculum-based measures (CBMs). However they did not benefit from accommodations on the computations, concepts, and applications CBMs. Teachers were found to over-award accommodations. Data-based rules supplemented teacher judgments in valuable ways.
Fuchs et al. (2000b)	Students with learning disabilities benefited differentially from the student read aloud accommodations, but not from large print or extended time accommodations. Teachers' decisions about appropriate accommodations did not correspond to the benefits students experienced. Data-based assessment techniques, based on differential gains from accommodations on several reading passages, predicted differential performance on large-scale assessments better than teacher judgments.
Helwig et al. (1999)	Students with low mathematics proficiency, regardless of reading ability, scored significantly higher when the video accommodation was provided. No significant association was found between the number of words, syllables, long words, or other language variables present in a test item and the difference in success rate on the standard or video version of the test. The performance of students with low reading fluency and above-average performance on the math skills test improved when the selected items were read aloud.
Johnson et al. (2000)	Scores of students with learning disabilities increased when math items were read to them. Reading the math items to non-learning disabled students had no effect on their performance.
Meloy et al. (2000)	The Read Aloud test administration condition benefited both LD-R and non LD students. LD-R and non LD students achieved higher scores on all tests under the read aloud conditions than under the standard administration condition. The mean scores for the non LD students were higher than those for the LD-R students in both conditions. The mean difference between conditions was larger (but not significantly significant) for the LD-R students.
Weston (1999)	Significant main effects were found for learning disabled and non-learning disabled students on both forms of the test (standard vs. oral presentation format). A larger effect from the accommodation was found for students with disabilities. Both types of math problems (calculation and word problems) showed an effect from oral presentation.

	Oral Presentation Had No Effect on Scores
Kosciolek & Ysseldyke (2000)	The students' test scores did not reveal a statistically significant interaction between the accommodation and the student group. The performance of students in each group was different; however, the differences were not statistically significant. The effect of the accommodation for students in general education was small and far from significant, while the effect on the special education group was larger and closer to statistical significance.
	Oral Presentation Altered Item Comparability
Bielinski et al. (2001)	The presence of the accommodation altered item difficulty beyond the effects accounted for by ability and disability status. Approximately half of the items on the reading test were considered to have differential item functioning when the accommodation was used, and one fifth of the items on the math test were considered to have differential item functioning when the read aloud accommodation was used.
Barton & Huynh (2000)	Results indicated associations between disability group and errors made, as well as level of achievement. Items characteristic of literal comprehension included in specific detail or reference questions, as opposed to inferential comprehension found in main idea questions, seemed to have the strongest association with disability status.
	Oral Presentation Did Not Alter Item Comparability
Pomplun & Omar (2000)	A two-factor model was considered adequate for all three groups of students. It appears that the same construct was being measured for students with learning disabilities who received the read aloud accommodation and students with learning disabilities who did not receive the read aloud accommodation.
	Extended Time Had Positive Effect on Scores
DiCerbo et al. (2001)	In general, participants scored 12 scale score points higher for divided-time administration than for the one-time administration. A time-by-reading comprehension ability interaction was also found. Poor and middle readers seemed to experience a greater benefit from the accommodation than higher readers.
Huesman & Frisbie (2000)	Little difference in the test performance of nonlearning disabled students, given directions to work at a normal rate, was found between the timing conditions (standard and extended). The test performance of nonlearning disabled students, given directions to take their time and work carefully, increased in the extended time condition. Similar results were found for learning disabled students in the sample; however, not to the same extent as students without learning disabilities. Students with learning disabilities made significantly greater gains on the ITBS Reading Comprehension Test under extended time than students without learning disabilities who received standard timing instructions. Non-learning disabled students given instructions to take their time did not perform any differently than students with learning disabilities who were given extra time.
Medina (2000)	Overall, students benefited from the extended test time. In general, the Pretest was a reasonable predictor of student performance on the Timed Test. Although for some individuals with learning disabilities extended time made a positive difference, students with learning disabilities did not benefit differentially from extended time.
Zuriff (2000)	An analysis of five studies failed to support the Maximum Potential Thesis that only students with learning disabilities benefit from extra examination time. In the studies examined, both students with and without disabilities benefited significantly from extra examination time. The research suggests that there may be subgroups of students with learning disabilities who benefit more from extra examination time.

	Extended Time Had No Significant Effect on Scores
Marquart (2000)	There were no significant differences in the performance of students with disabilities, students academically at-risk, and students functioning at or above grade level on the standard versus the extended time conditions. There was no significant difference in the effect sizes for the accommodation among the three student groups. Student reactions to the test were more positive when they were given extra time rather than the standard time to complete the test.
Schulte et al. (2001)	Students with and without disabilities benefited from test accommodations. Students with disabilities experienced a larger effect in the accommodated condition than students without disabilities. However, the difference between groups was not statistically significant for total scaled scores. Students with disabilities benefited more than students without disabilities on the multiple-choice items, but not on the constructed response items. The accommodation package of extra time and read aloud did not have a differential impact for students with disabilities when compared to students without disabilities. Students with disabilities receiving accommodation packages other than extra time and read test items to students experienced a significant and differential impact of testing accommodations on math scores.
Walz et al. (2000)	For special education students, the mean number correct was similar for both the Multiple Day and One Day conditions. The general education students did not perform as well when taking the test across multiple days as they did when taking it on one day.
	Student-Paced Video Accommodation Had Positive Effect On Scores
Hollenbeck et al. (2000)	The mean scores for students with disabilities and the lowest general education students were slightly raised when the accommodation was student-paced versus when the teacher paced the accommodation on the video.
	Examiner Familiarity Had Positive Effect on Scores
Szarko (2000)	A significant difference between groups on the Cognitive Verbal and Performance subscales of the PEP-R was observed. A significant difference in stereotypic behaviors was also observed between the two groups. Examiner familiarity had significant positive effects on the behavior and testing performance of children with autism.
	Type of Calculator Had No Significant Effect on Scores
Hanson et al. (2001)	No significant differences in accuracy, timing, calculator use, calculator difficulties, or number of keystrokes used were uncovered among the different calculator type conditions. Background characteristics (sex, family income, calculator complexity, ethnicity, and difficulty level of the student's last math class) were not related to differences in performance by calculator type. Calculator preference depended on how complex the student's calculator was in comparison to the standard calculator.
	Simplified Language Had No Significant Effect on Scores
Tindal et al. (2000)	Simplifying the language did not affect the test. Students with and without disabilities performed equally well in either simplified or standard conditions. The implication is that simplifying language can be done without changing the construct.
	Sign Language Accommodation Altered Item Comparability
Johnson et al. (2001)	The results suggest that the use of sign language as an accommodation presents political, practical, and psychometric challenges. The data showed that sign language translation may result in the omission of information required to answer a test item correctly.

	Meta Analyses Showed Positive Effects Of Accommodation Use On Scores
Chiu & Pearson (1999)	The meta analysis found that compared to conditions without accommodations, special education (SE) and limited English proficient (LEP) students who received accommodations increased their scores by an average of .16 standard deviation. In comparison to regular education students, special education students and limited English proficient students exhibited an average accommodation advantage of .10 standard deviation. Elementary as well postsecondary students benefited from accommodations.
Elliott et al. (2001)	More than 75 percent of the testing accommodation packages, suggested by students' IEP teams, had a moderate to large effect on their test scores. Testing accommodations had a positive impact on scores of students without disabilities. The effects of recommended accommodations were not positive for a small number of students.
McKevitt (2000)	Accommodations had a medium to large positive effect for 78.1% of students with disabilities and 54.5% of students without disabilities. Small or zero effects were found for 9.6% of students with disabilities, and 32.3% for students without disabilities. Accommodations had negative effects on the scores of 12% of students with disabilities and 13.1% of students without disabilities. Generally, the resulting individual effect size when comparing accommodated scores to non-accommodated scores was .88 for students with disabilities, .44 for students without disabilities who received the standard package, and .45 for students without disabilities received accommodations recommended by their teachers. The performance of students with disabilities under the accommodated condition was slightly lower than students without disabilities, while the performance of students with disabilities without accommodations was well below the average performance of other students.
McKevitt et al. (2000)	Accommodations had a medium to large positive effect on 81% of students with disabilities and 51% of students without disabilities. Small to zero effects were found for 5% of students with disabilities, and 41% of students without disabilities. Accommodations had a negative effect on the scores of 14% of students with disabilities and 7.8% of students without disabilities. Students without disabilities experienced a slight increase in scores when compared to those who received the standard package. When comparing accommodated to nonaccommodated scores the resulting effect size was .94 for students with disabilities, .44 for students without disabilities receiving the standard accommodation package, and .55 for students without disabilities receiving the teacher recommended accommodations.
McKevitt et al. (1999)	Overall, testing accommodations had medium to large positive effects for 75.6% of students with disabilities and 55.9% of students without disabilities. Small to zero effects were found for 7.3% of students with disabilities and 38.2% of students without disabilities. Accommodations had a negative effect on the scores of 17.1% of students with disabilities and 5.9% of students without disabilities. The average effect size when comparing accommodated scores to non-accommodated scores was .83 for students with disabilities.
	Some Items Exhibited DIF Under Accommodated Conditions
Lewis et al. (1999)	Only a few items exhibited differential item functioning (DIF) for participants under the accommodated conditions. More English language arts items than math items exhibited DIF. The number of DIF items ranged from 1-3 for math items and 1-5 for language arts items depending on the disability group in question.
Zurcher (1999)	Associations between grade point average and test scores were not weaker for tests with accommodations. Instead, they were comparable to scores of non-learning disabled students who received standard administrations of the test.

	Some Items Exhibited DIF Under Accommodated Conditions
Elliott et al. (1999)	Students in special education scored lower than students in general education who did not receive accommodations, regardless of the number of accommodations used. Differences in findings within content areas were found; however, the overall findings were that accommodations did not inflate test scores. Higher performance levels were found in the extended time condition than in other accommodated conditions. In some instances, the factor structures of tests appeared different with different accommodations.
Koretz & Hamilton (1999)	Accommodation effects were stronger on open-response test items than on multiple-choice items across all grades. In most cases, the differences in scores between disabled and non-disabled students were larger on multiple-choice components in the early grades, and similar or greater differences for open-response components in the higher grades. Correlations among parts of the assessment were different for accommodated students with disabilities than for other students, with higher correlations across subjects for the open-response items.
Koretz & Hamilton (2001)	Accommodations were used liberally, with extra time and testing in a separate location being provided the most. Students with disabilities scored approximately two thirds to one and one third standard deviations below other students, and the responses of a high percentage of students with disabilities to open-response items were weak or unscorable.
Koretz & Hamilton (2000)	The results suggest that most students were included in the assessment; however, the inappropriate use of accommodations may lead to unreliable scores. In particular, it seemed students receiving the dictation accommodation obtained improbably high scores. For open response item formats, item-test correlations were no different for students with disabilities (with or without accommodations) than for students without disabilities, and differential item functioning was relatively infrequent in some subject areas. There were several cases of notable differential item functioning shown by students receiving accommodations, mainly in mathematics.
	Accommodation Decisions are Based on Educator Beliefs
Bourke et al. (2000)	Beliefs about the helpfulness and need for instructional accommodations were associated with the provision of accommodations. Perception of support from the University influenced the ease of providing instructional accommodations.
Schulte et al. (2000)	Educators did not recommend using more accommodations for students who have severe disabilities than for students who have more mild disabilities. Educators recommended more accommodations be used on a performance assessment task than on a multiple-choice task. Results suggest that educators perceive the Assessment Accommodation Checklist (ACC) as having good content validity and believe it is useful for generating ideas and documenting accommodations used for students with disabilities.
Simpson et al. (1999)	More than 86% of teachers surveyed recommended assessment participation for students with Asperger syndrome, but of teachers of students with moderate to severe autism, only 54.9% and 8.33% respectively, believed that students should participate.
Vogel et al. (1999)	Faculty members were slightly more willing to offer teaching accommodations than examination accommodations. For instructional accommodations, faculty members seemed most willing to allow students to tape-record lectures and least willing to provide supplementary materials or assignments in alternative forms. For examination accommodations, faculty members seemed most willing to allow extended time for exams to be proctored in the office of support services for students with disabilities. Faculty members seemed least willing to change exam formats.

Appendix H

Summary of Limitations Cited by Researchers*

Authors (year)	Unknown Variations Between Students
Bourke et al. (2000)	The generalizability of the results from this study is limited due to the fact that the survey was administered to participants at only one university. The participants may have been self-selecting since only one-third of the possible respondents returned completed surveys.
Helwig et al. (2000)	It is possible that writing designed for different purposes requires students to utilize different schemata, resources, or skills. Thus, the wide range of students and varied conditions used in this study are possible limitations of this research.
Hollenbeck et al. (2000)	The study was confounded by medium. A teacher-paced accommodation defined solely by video technology or a student-paced accommodation by computer technology is inappropriate. Technology was used as the vehicle for providing the accommodation. While, technology is neutral, the accommodation connected with that technology appears to be person dependent. In addition, each participant's proximal distance from the screen may be a confound. Individual students may respond better to either the video or computer accommodation based on the screen distance they prefer.
Hollenbeck et al. (1999a)	While students' perceptions of their keyboarding skills were measured, no base keyboarding score was mandated for inclusion in the study. Teacher skills may also have influenced the results. Observations revealed that teachers for some groups appeared to teach the writing process more intensely using computers specific to intervention than did the teachers in other groups.
Huesman & Frisbie (2000)	The lack of information on the type of disability, the severity of the disability, and the presence of multiple disabilities for the LD group confounds interpretation of the observed results and limits the generalizability of the results to elementary-grade students with a primary label of learning disability. In addition, the use of school-identified samples of students with learning disabilities may be biased in terms of gender. If females are underrepresented or if the females identified have more or less severe deficits than the majority of LD females, the results may not be generalizable to the general elementary-grade LD population.
Lewis et al. (1999)	The number of different special needs/accommodated combinations that could be analyzed was limited by the diversity of special needs students in the sample identified as taking the test under accommodated conditions.
Schulte (2000a)	The vignettes used in this study lacked adequate information about the students that was needed to make accurate recommendations and/or ratings. The density of the materials used in this study may have confused some participants and contributed to the low return rate of surveys through the mail.
Schulte et al. (2001)	Specific accommodations were not studied in isolation. Consequently, for students receiving a package of accommodations, it could not be determined whether specifically just one or two accommodations made the difference or whether the whole package interacts in such a way as to remove barriers from performance. Only one subject (mathematics) was targeted, thus researchers could not determine whether the accommodations have a different impact in other subject areas. IEP teams or teachers recommended the accommodations for each student; however, the students did not provide input into the accommodations they thought they would need. Consequently, it is possible that the accommodations were not totally matched to individual needs or preferences.

Authors (year)	Unknown Variations Between Students
Walz et al. (2000)	During one of the testing sessions a temporary disruption occurred affecting only the Special Education group included in the study. The extent to which the disruption may have impacted the students' test performance on that particular day is unknown. Other accommodations that may have been provided under normal circumstances were not provided in this study, and thus the real world validity may have been compromised.
Weston (1999)	For non-disabled students the evidence may be flawed by the following methodological problems: (1) very low readers in the regular classroom did not seem to profit from the accommodation, although this result may be due to their poor representation in the sample, (2) item content did not seem to make a difference on the effects of the accommodation, and (3) these students performed better on a number of items in the paper and pencil format.
Zuriff (2000)	The author noted several limitations of some or all the studies reviewed in this analysis: small sample size or lack of statistical support, utilization of the no-review policy, which has serious methodological flaws, significant overlap between the performances and improvements of the participants with learning disabilities and the nonlearning disabled participants, a lack of individual data, which makes it impossible to exclude the possibility that there is an identifiable subgroup of students with learning disabilities who benefit from extra time even when the nonlearning disabled students as a group do not show significant improvements.
	Small Sample Size
Elliott et al. (1999)	More research needs to be done to determine when accommodations are used since small numbers of students were involved in some of the analyses in this study.
Johnson (2000)	The importance of the assessment tool used, the Washington Assessment of Student Learning (WASL) caused significant anxiety for the students participating in the study. The anxiety surrounding the test contributed to several limitations of this study: a change in the way the test was presented to students, the small sample size that contributed to the limited power of the study to detect possible significant differences among scores, and the lack of a special education control group that made it difficult to compare students with learning disabilities to a control group.
Koretz & Hamilton (2001)	The small number of students with disabilities hindered the analysis of data. The limitation of small sample size was compounded by the use of nonequivalent forms and the heterogeneity of students with disabilities. Only very large discrepancies in performance can be statistically significant due to the small samples.
Kosciolek & Ysseldyke (2000)	The sample size of the groups was relatively small making it difficult to determine a statistically significant difference in the effect of the accommodation for students by group.
Russell (1999)	A small group of students from one urban district was included in the study. This may be problematic because there may be differences in the way schools and students use computers that can lead to different effects for students in different settings. Tests were not administered under formal, controlled testing conditions and this may have led to increased distractions and affected student performance and motivation. Student performance may also have been affected by the time limits of the tests. It was impossible to estimate the effect of taking open-ended tests on computer for students who are proficient or advanced keyboarders because the sample of students included in the study had relatively slow keyboarding skills.

	Small Sample Size
Russell & Plati (2000)	It was not possible to examine the mode of administration effect across the full range of performance levels because very few students included in the sample performed at low levels on the composition items. The students' familiarity with computers and high level of keyboarding speed made it difficult to examine the mode of administration effect for both students with low levels of keyboarding and students not accustomed to computers. Furthermore, there were several reasons why the data collected may not accurately reflect actual levels of student performance. These include: the composition items did not get factored into grades, the exam may not have been proctored by a teacher in the school, the students in the computer group finished their compositions faster than students in the paper groups, and a few responses generated on the computer failed to address the question posed.
Simpson et al. (1999)	The vignettes in this study were used as a basis for describing students. As such, these hypothetical descriptions may not have depicted students with autism spectrum disorders in the manner in which teachers perceived actual children. The specific accommodations recommended for the student described as having severe autism were based on a very small sample.
Tindal et al. (2000)	Several explanations are offered as to why the findings from this study are anomalous and would change with further research. For example, too few students were included in the study, the researchers incorrectly recommended against the use of language simplification as an effective accommodation, and the process used to simplify the language on the test may have been directed at the wrong variables.
Zuriff (2000)	The author noted several limitations of some or all the studies reviewed in this analysis: small sample size or lack of statistical support, utilization of the no-review policy, which has serious methodological flaws, significant overlap between the performances and improvements of the participants with learning disabilities and the nonlearning disabled participants, a lack of individual data, which makes it impossible to exclude the possibility that there is an identifiable subgroup of students with learning disabilities who benefit from extra time even when the nonlearning disabled students as a group do not show significant improvements.
	Nonstandard Administration Across Proctors and Schools
Bielinski et al. (2000)	The administration of the read aloud accommodation might have been flawed causing the read aloud accommodation to result in more differential item functioning (DIF) items. The read aloud accommodation could have differed from proctor to proctor since it was not administered uniformly across schools. Group presentation of the read aloud accommodation might have influenced student performance by making them feel uncomfortable in the presence of peers. The analysis only included students with a reading disability. However, it is possible that these students might not actually be the students who actually require the read aloud accommodation.

	Nonstandard Administration Across Proctors and Schools
Hollenbeck, et al. (2000)	The study was confounded by medium. A teacher-paced accommodation defined solely by video technology or a student-paced accommodation by computer technology is inappropriate. Technology was used as the vehicle for providing the accommodation. While, technology is neutral, the accommodation connected with that technology appears to be person dependent. In addition, each participant's proximal distance from the screen may be a confound. Individual students may respond better to either the video or computer accommodation based on the screen distance they prefer.
Hollenbeck et al. (1999a)	While students' perceptions of their keyboarding skills were measured, no base keyboarding score was mandated for inclusion in the study. Teacher skills may also have influenced the results. Observations revealed that teachers for some groups appeared to teach the writing process more intensely using computers specific to intervention than did the teachers in other groups.
Koretz & Hamilton (2000)	The Kentucky Department of Education (KDE) did not collect information on which accommodations were used on which format, making it difficult to determine whether differences in relationships between accommodations use and performance across formats resulted from differences on the actual effects of the accommodations or whether those accommodations were simply used to varying degrees across formats.
Russell (1999)	A small group of students from one urban district was included in the study. This may be problematic because there may be differences in the way schools and students use computers that can lead to different effects for students in different settings. Tests were not administered under formal, controlled testing conditions and this may have lead to increased distractions and affected student performance and motivation. Student performance may also have been affected by the time limits of the tests. It was impossible to estimate the effect of taking open-ended tests on computer for students who are proficient or advanced keyboarders because the sample of students included in the study had relatively slow keyboarding skills.
Walz et al. (2000)	During one of the testing sessions a temporary disruption occurred affecting only the Special Education group included in the study. The extent to which the disruption may have impacted the students' test performance on that particular day is unknown. Other accommodations that may have been provided under normal circumstances were not provided in this study, and thus the real world validity may have been compromised.

*Some researchers cited limitations in more than one category.

Appendix I

Summary Recommendations by Researchers

Authors (year)	Replicate Results for Validation and Generalization
Chiu & Pearson (1999)	There is a need for more studies that allow us to understand and predict the effects of various presentation formats and response formats in relation to particular target populations.
Fuchs et al. (2000a)	Additional research might productively focus on the development of similar DATA-based methods for determining accommodations on reading tests or the exploration of other demographic markers associated with differential accommodation boosts for students with learning disabilities.
Helwig et al. (1999)	Follow-up studies are needed to help identify types of word problems best suited to video accommodations, similar types of accommodations that are most effective, and students who would benefit from alterations in test format.
Hollenbeck et al. (2000)	In the future researchers should continue to work toward a generalized or global validation of testing accommodations, consider pacing as it relates to specific accommodations, and examine the correlation between accommodations and the student's degree of conceptual understanding.
Johnson (2000)	Continued research is needed to examine the effects of reading a test aloud in content areas other than math for all grade levels.
Koretz & Hamilton (2001)	Future research should monitor examinations for several years in order to estimate the participation rates of students with disabilities, and then set a target for participation rates based on the characteristics and uses of the examinations and the nature of alternatives available to students. Additional exploration is needed to help evaluate the appropriateness of the current use of accommodations. There should be further exploration of possible differences in the difficulty of different types of items and forms included in the Regents English test for students with special needs, including both students with disabilities and English language learners.
Lewis et al. (1999)	Further research is needed to establish the comparability of scores for students with special needs who take the ISTEP + Graduation Qualifying Exam (GQE) under accommodated conditions and regular education students who take the test under nonaccommodated conditions.
Meloy et al. (2000)	There is a need for more studies that incorporate different pacing arrangements and different expressive styles of reading. Varying approaches concerning pacing could be examined, such as single readings of questions/passages versus two readings and/or varying times given between question stem and response options.
Pomplun & Omar (2000)	More research is needed to confirm the findings from this study because it was one of the first studies to use covariance and mean structures to examine the comparability of test scores of students with learning disabilities and general education students. Research on differential item functioning would be an important addition to knowledge about the effects of accommodations on score comparability. Further research is needed to investigate the nature of accommodations in more depth.
Vogel et al. (1999)	There is a need to replicate the findings of this study in similar universities and contrasting institutions and in different regions of the country to determine whether these findings generalize to other institutions. Other strategies should be utilized to cross-validate the findings from this study, such as asking students with LD and service providers about faculty practices (rather than attitudes) and surveying administrators who see to student appeals and policies regarding course substitutions, waivers of a specific course, or exemptions from taking a specific examination or part of an examination for licensure.

Authors (year)	Replicate Results for Validation and Generalization
Walz et al. (2000)	Additional multiple day testing accommodation studies would be beneficial with different types and lengths of tests that include wider age ranges of students. Researchers should consider how groups or pairings of accommodations, which may normally be used by students, are dealt with when attempting to study a specific accommodation. There is a need for increased knowledge and awareness of language-related factors and other non-language factors among students, as well as closer collaboration between schools and researchers.
	Investigate Specific Disability Associations
Barton et al. (2000)	In order to reveal a clearer description of the types of items on which specific disability groups make errors, future analyses will need to investigate specific disability associations for strongly associating items with their content. Additionally, an analysis of non-accommodated examinees might offer further insight into common reading patterns.
Calhoon et al. (2000)	Studies are needed to investigate whether a reading test accommodation is more effective for students with LD at different reading levels. It is also important to know if learning disabled students react differentially to the provision of a reader in different content areas. Similarly, the effect of a human versus a computer needs investigation. More research is needed on the efficiency of a human versus computer reader, savings in time and monies, as well as students' perception about the different types of reading test accommodations. Additionally, an investigation of the effect of video as a test accommodation should be carried out with a focus on training students with learning disabilities to extract information from video and to transfer knowledge to test-taking situations. Other important issues for future research are the differing effects of test accommodations for students with and without learning disabilities.
Elliott et al. (2001)	Future work on testing accommodations will be improved if the cognitive ability of student participants is controlled, or treated like a covariate. There are still important questions that need to be answered concerning the point at which change from standard test administration actually changes the task being measured and regarding the issue of determining when the tasks and resulting scores are no longer comparable and valid.
Schulte et al. (2000a)	Future researchers should consider the role that severity of the disability could play into number of accommodations required and educators' perceptions of those accommodations. Researchers may also consider collecting more information from teachers regarding the feasibility of the accommodations on the Assessment Accommodation Checklist (AAC) given the amount of resources available to them. The most commonly used or recommended accommodations need to be investigated empirically to determine more objectively whether they might invalidate an assessment.
	Conduct More Detailed Non-experimental Studies to Provide Richer Data
Koretz & Hamilton (1999)	Descriptive studies of the performance of students with disabilities in large-scale assessments are needed. In addition, more detailed non-experimental studies that would require richer data than can be obtained through routine data collection offer a much more descriptive view of assessment and accommodations and a stronger basis for hypothesizing about the effects of format, accommodations, and other factors.
Schulte et al. (2001)	Continue investigating the effect of accommodations on constructs in math and other areas, consider including students in the development of accommodation plans, and look more closely at the impact of testing accommodations on multiple-choice versus constructed-response types of items. Future research could study accommodations in isolation and in combination to examine effects and examine how extra time impacts student performance.

	Conduct More Detailed Non-experimental Studies to Provide Richer Data
Zuriff (2000)	Future research should examine solutions to the modified test integrity problem other than the Maximum Potential Thesis (MPT). Given the variability in the effects of extra examination time on the scores of college students with and without learning disabilities, research should concentrate on determining the conditions under which students benefit from extra examination time. Researchers should look more cautiously at the rationale for test accommodations.
	Increase Researcher Control of Testing Process
Huesman & Frisbie (2000)	Future investigations of the score comparability of students with and without learning disabilities under extended time conditions should involve more researcher control of the entire testing process. Procedures such as distributing special directions, conducting workshops for test administrators on appropriate special test administration procedures, and using multiple trained observers in the classrooms during testing would help ensure uniform test administration and improved data collection across buildings and consequently increase the power of the statistical tests and make interpretation of the results more clear and generalizable.
Koretz & Hamilton (2000)	Better data collection systems are needed to provide more specific information on how accommodations are selected and used. Particularly, data on the kinds of accommodations used in everyday instruction would provide important evidence about the appropriateness and effectiveness of assessment accommodations.
	Study Larger Sample
Burk (1999)	Tests should be administered to other disability populations as well as larger numbers of students with learning disabilities. Other tests should be adapted, and more thorough research is needed as to the effects of particular accommodations on specific students as IEP teams recommend them. Further research is needed using unabbreviated "real-world" tests to determine whether fatigue and/or anxiety would become factors in a real world setting.

The College of Education
& Human Development

UNIVERSITY OF MINNESOTA

NCEO is an affiliated center of the Institute on Community Integration

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").